

Thu May 8 16:14:00 2003

us-09-658-621b-2.rat

Page 1

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

CM protein - protein search, using bw model

Run on: May 1, 2003, 23:09:26 ; Search time 25.5357 Seconds
(Without alignments)
570.352 Million cell updates/sec

Title: US-09-658-621B-2

Sequence: 1 MPTGTQSPFFLLLTVLTV.....NGSSSLSTYTPAVATSNL 495

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents AA: *
2: /cgn2_6/prodata/1/aa/5A.COMB.pap: *
3: /cgn2_6/prodata/1/aa/5B.COMB.pap: *
4: /cgn2_6/prodata/1/aa/6A.COMB.pap: *
5: /cgn2_6/prodata/1/aa/6B.COMB.pap: *
6: /cgn2_6/prodata/1/aa/BACKFILES1.pap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 2019 | 78.8 | 2035 | 2 | US-08-479-537A-2 |
| 2 | 2019 | 78.8 | 2035 | 4 | US-09-083-116-2 |
| 3 | 2019 | 78.8 | 2035 | 4 | US-09-134-916A-2 |
| 4 | 1091 | 42.6 | 1867 | 2 | US-08-479-537A-5 |
| 5 | 1091 | 42.6 | 1867 | 4 | US-09-083-116-5 |
| 6 | 234.5 | 9.2 | 907 | 3 | US-09-134-916A-5 |
| 7 | 234.5 | 9.2 | 907 | 4 | US-08-783-774-2 |
| 8 | 234.5 | 9.2 | 907 | 5 | US-09-328-559A-1 |
| 9 | 234.5 | 9.2 | 907 | 5 | PCT-US95-04611A-19 |
| 10 | 228 | 8.9 | 978 | 4 | US-09-556-706B-2 |
| 11 | 207.5 | 8.0 | 750 | 4 | US-08-053-29A-4 |
| 12 | 205 | 8.0 | 40 | 2 | US-08-288-059-1 |
| 13 | 205 | 8.0 | 40 | 2 | US-08-288-059-1 |
| 14 | 205 | 8.0 | 1060 | 4 | US-08-911-193-2 |
| 15 | 198 | 7.7 | 2137 | 4 | US-09-134-001C-4463 |
| 16 | 192 | 7.5 | 941 | 4 | US-07-757-022B-14 |
| 17 | 192 | 7.5 | 1022 | 4 | US-07-757-022B-84 |
| 18 | 192 | 7.5 | 1038 | 4 | US-07-757-022B-74 |
| 19 | 192 | 7.5 | 1049 | 4 | US-07-757-022B-58 |
| 20 | 192 | 7.5 | 1140 | 4 | US-07-757-022B-104 |
| 21 | 192 | 7.5 | 1270 | 4 | US-07-757-022B-44 |
| 22 | 192 | 7.5 | 1311 | 4 | US-07-757-022B-42 |
| 23 | 192 | 7.5 | 1311 | 4 | US-07-757-022B-142 |
| 24 | 192 | 7.5 | 1314 | 4 | US-07-757-022B-50 |
| 25 | 192 | 7.5 | 1320 | 4 | US-07-757-022B-46 |
| 26 | 192 | 7.5 | 1320 | 4 | US-07-757-022B-60 |
| 27 | 192 | 7.5 | 1354 | 4 | US-07-757-022B-48 |

| | | | | | |
|----|-------|-----|------|---|--------------------|
| 28 | 192 | 7.5 | 1361 | 4 | US-07-757-022B-40 |
| 29 | 192 | 7.5 | 1363 | 4 | US-07-757-022B-52 |
| 30 | 192 | 7.5 | 1404 | 4 | US-07-757-022B-2 |
| 31 | 192 | 7.5 | 1404 | 4 | US-07-757-022B-62 |
| 32 | 189 | 7.4 | 805 | 4 | US-09-103-429A-4 |
| 33 | 188.5 | 7.4 | 786 | 4 | US-09-513-429A-3 |
| 34 | 174 | 6.8 | 1125 | 4 | US-09-513-783A-152 |
| 35 | 174 | 6.8 | 1185 | 4 | US-09-041-866-23 |
| 36 | 174 | 6.8 | 1610 | 4 | US-09-513-783A-22 |
| 37 | 167.5 | 6.5 | 2972 | 4 | US-09-579-181-2 |
| 38 | 167.5 | 6.5 | 3118 | 4 | US-09-579-181-1 |
| 39 | 165 | 6.4 | 401 | 6 | 5252556-1 |
| 40 | 162.5 | 6.3 | 1581 | 4 | US-09-110-517-2 |
| 41 | 162 | 6.3 | 267 | 4 | US-08-818-112-142 |
| 42 | 162 | 6.3 | 267 | 4 | US-08-818-111-137 |
| 43 | 162 | 6.3 | 267 | 4 | US-09-056-556-142 |
| 44 | 162 | 6.3 | 267 | 4 | US-09-072-596-137 |
| 45 | 161.5 | 6.3 | 1719 | 2 | US-08-459-568-4 |

ALIGNMENTS

RESULT 1
US-08-479-537A-2
Sequence 2, Application US/08479537A
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIRBY, Marie-Paule
APPLICANT: LATHÉ, Richard
APPLICANT: HAREUYENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DONNE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Thu May. 8 16:14:00 2003

us-09-658-621b-2.rat

Page 3

FEATURE: Peptide
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
FEATURE: or CCG, and Ala = GCT, GCC, GCA, or GCG."
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2
Query Match 78.8%; Score 2019; DB 4; Length 2035;
Best Local Similarity 85.3%; Pred. No. 9.7e-146;
Matches 399; Conservative 5; Mismatches 62; Indels 2; Gaps 1;
QY 28 SSTPGKEKTSATORSSVPSSTERNKAVMTSSVLSHSPGSGSTTQGDVTLAPATERA 87
DB 1570 STAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHG--VTSAPDXRFX 1627
QY 88 SGAATWGDVTSVPTSPALGSTTPPAHDVTSAPDNKPAVPGSTAPXAHGVTSAPPTRP 147
DB 1628 PSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRFX 1687
QY 148 PSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRFX 207
DB 1688 PSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRFX 1747
QY 208 SASIVHNGTSABATTPRASKSTPSIPSHSDPTTLASHTKIDASTHSTVPLTS 267
DB 1748 SASIVHNGTSABATTPRASKSTPSIPSHSDPTTLASHTKIDASTHSTVPLTS 1807
QY 268 SNHSTSPQLSTGVSPFSLFSLHSLDPSIDYQELGROISMPFOIKKGGFL 327
DB 1808 SNHSTSPQLSTGVSPFSLFSLHSLDPSIDYQELGROISMPFOIKKGGFL 1867
QY 328 GLSNKKPPGVSVOVTLAFREGTINVHDETOPFOVYKTAASRYVLTISVSVSHVPP 387
DB 1868 GLSNKKPPGVSVOVTLAFREGTINVHDETOPFOVYKTAASRYVLTISVSVSHVPP 1927
QY 388 FSNAGAGVPGMGIALVLCVLAIVYLALAVCCCRKRVGQDIFPARDTYHMS 447
DB 1928 FSNAGAGVPGMGIALVLCVLAIVYLALAVCCCRKRVGQDIFPARDTYHMS 1987
QY 448 EYPTHTHGRVPPSSSTRSPYKVSAGNGGSLSTYNPAVAATSNL 495
DB 1988 EYPTHTHGRVPPSSSTRSPYKVSAGNGGSLSTYNPAVAATSNL 2035
RESULT 3
US-09-134-916A-2
Sequence 2, Application US/09134916A
GENERAL INFORMATION:
PATENT NO. 6328356
APPLICANT: CHIRON, Pierre
APPLICANT: CHIRON, Pierre
APPLICANT: KERN, Marie-Pierre
APPLICANT: LAMPE, Richard
APPLICANT: HARIVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE: 07-JUN-1995
CLASSIFICATION: US/08/479,537
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Tebkin, Robin L.
REGISTRATION NUMBER: 35,017753-025
REFERENCE/AGENT INFORMATION:
TELEPHONE: (703) 816-6620
TELEFAX: (703) 816-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACU
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2
Query Match 78.8%; Score 2019; DB 4; Length 2035;
Best Local Similarity 85.3%; Pred. No. 9.7e-146;
Matches 399; Conservative 5; Mismatches 62; Indels 2; Gaps 1;
QY 28 SSTPGKEKTSATORSSVPSSTERNKAVMTSSVLSHSPGSGSTTQGDVTLAPATERA 87
DB 1570 STAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHG--VTSAPDXRFX 1627

Thu May 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 4

| | | | |
|----|------|---|-------|
| QY | 88 | SOSAAITWCOODVTSVPTRALGASTTTPHADVYSAEDNKPAGSGTAPPAHGVYSAEDTRRP | 1.47 |
| Db | 1628 | PGSTAPKXHGVTAPKPRXKPGSTAPPAHGVTSADNKPXGSGTAPXAGVYSAEDTRRP | 1.687 |
| QY | 148 | PGSTAPPAHGVTSAPTRPAPGSTAPPAHGVTSADNKPALGSGTAPPVYNTVTSAGSAG | 2.07 |
| Db | 1688 | PGSTAPPAHGVTSAPKPRXKPGSTAPPAHGVTSADNKPXGSGTAPPVYNTVTSAGSAG | 1.744 |
| QY | 208 | SASITVHNKTSARATTPKASSTPPSI.PSHHSIDPTTLASHSTKTDASTHASTVPLETS | 2.67 |
| Db | 1748 | SMSTLVHNSARATTPKASSTPPSI.PSHHSIDPTTLASHSTKTDASTHASTVPLETS | 1.807 |
| QY | 268 | SNHSTSPGLSTGVSPEFLSPHSMLOKNSLAPBEDTVQELORISLSEMPATYKQKQGL | 3.27 |
| Db | 1808 | SNHSTSPGLSTGVSPEFLSPHSMLOKNSLAPBEDTVQELORISLSEMPATYKQKQGL | 1.867 |
| QY | 328 | GLSNTKRPSPSVVOLLTA.PEFGTINVHDEVOPNOXYTEASRYVLTISGVSSHPP | 3.87 |
| Db | 1868 | GLSNTKRPSPSVVOLLTA.PEFGTINVHDEVOPNOXYTEASRYVLTISGVSSHPP | 1.927 |
| QY | 388 | PSAGSGAGVGOMGIALVLCVLAALAYLLTAALAVCCCRKRYGGLDIPARDTYHMS | 4.47 |
| Db | 1928 | PSAGSGAGVGOMGIALVLCVLAALAYLLTAALAVCCCRKRYGGLDIPARDTYHMS | 1.987 |
| QY | 448 | EYPTVHTHGRVYPPSGTDRSPYEKVSANGGSSLSYTNPPVAATSNL | 4.95 |
| Db | 1988 | EYPTVHTHGRVYPPSGTDRSPYEKVSANGGSSLSYTNPPVAATSNL | 2035 |

US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENY, Mera
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUDNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: D1A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO CQT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teekin, Robin L.
REGISTRATION NUMBER: 35, 030
REFERENCE/DOCKET NUMBER: 017753-025

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
STRATEGY: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
/note= "The amino acids spanning
128 to 1727 constitute a repeated region wherein the repea
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
/note= "Amino acid 134 is Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, C
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
/note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
/note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
/note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."

```

```

QY      28  SGPAGEKETISATORSSVPSTSKXNAVMTSSVLSSHSPPGSGSSTTQOQDVTLAPTEPA  87
Db      1570  STRPFAHGVTSADDXRPPGSTAPXAHGVTSADDXRPPGSTAPXAHG--TTSAPDXRPX  1627
                                     42.6%; Score 1091; DB 2; Length 1667;
                                     95.5%; Ident 100; Mismatches 61; Indels
                                     5; Gaps
QY      88  SSGAAITWCQDVTSVPTREPALGSGTTPPAHDTVSAPDNKPAFGSTAPXAHGVTSAPDXRPP  147
Db      1628  FGSITAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPP  1687
QY      148  FGSITAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDNKPAFGSTAPXAHGVTSAPDXRPP  207
Db      1628  FGSITAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPP  1747
QY      168  FGSITAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPP  267
Db      168  FGSITAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPPGSTAPXAHGVTSAPDXRPP  1747
QY      1748  SAGTLYNNQTSAAITTPPAKSPISLSHSDPTPLTASSIKTDASHSTHSTVPLTS  1897
Db      1748  SAGTLYNNQTSAAITTPPAKSPISLSHSDPTPLTASSIKTDASHSTHSTVPLTS  1897
QY      248  SNHSTSPOLSTGVSFFFLSGSHINQFNPSLDEPSTDYQLOLRDISM  316
Db      1808  SNHSTSPOLSTGVSFFFLSGSHINQFNPSLDEPSTDYQLOLRDISM  1856

RESULT 5
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
APPLICANT: CHAMBERON, Pierre

```


Thu May. 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 5

APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUENT, Marie
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE: 08/03/1993
PRIORITY INFORMATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MODEL TYPE: peptide
NAME/KEY: Peptide
LOCATION: 128, 1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5
Query Match 42.6%; Score 1091; DB 4; Length 1867;
Best Local Similarity 76.5%; Pred. No. 8e-75;
Matches 221; Conservative 5; Mismatches 61; Indels 2; Gaps 1;
QY 28 SPTGGEXETATQGRSSVSTENQAVSMSTSVLSHSPSGSSTTQODVTLAPTEPA 87
DB 1570 STAPAHGVTSAPDXRXPXGTAAPAHGVTSAPDXRXPXGTAAPAHG--VTAPDXRXP 1627
QY 88 SGAATWQODVTSVPTRPALGSTPPADVTSAPDNKPPAGSTAPPAHGVTSAPDTRP 147
DB 1628 PGSTAPXAHGVTSAPDXRXPXGTAAPAHGVTSAPDXRXPXGTAAPXAHGVTSAPDXRXP 1687
QY 148 PGSTAPAHGVTSAPDTRPAPGTAAPAHGVTSAPDNKPPAGSTAPPAHGVTSAPGASG 207
DB 1688 PGSTAPXAHGVTSAPDXRXPXGTAAPAHGVTSAPDXRXPXGTAAPAHGVTSAPGASG 1747
QY 208 SASTLVNNGTSARATTPASKSTPSPISPHSGDTPTTLASHGKTDASTHGVPLTS 267
DB 1748 SASTLVNNGTSARATTPASKSTPSPISPHSGDTPTTLASHGKTDASTHGVPLTS 1807
QY 268 SMTSTSPQSTGVSPFPLSFHISMTLQPNLSLEDPETDYQELORISFM 316
DB 1808 SMTSTSPQSTGVSPFPLSFHISMTLQPNLSLEDPETDYQELORISFM 1856
RESULT 6
US-09-134-916A-5
Sequence 5, Application US/09134916A
Patent No. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUENT, Marie
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY INFORMATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576

Thu May 8 16:14:00 2003

us-09-658-621b-2.rat

Page 6

```

FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDNESS: single
MOLWT: 206,000
MOLTYPE: linear
FEATURES:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-5

Query Match 42.6%; Score 1091; DB 4; Length 1867;
Best Local Similarity 76.5%; Pred. No. 8e-75;
Matches 221; Conservative 5; Mismatches 61; Indels 2; Gaps 1;

QY 28 SSTPGGKERTATQSSVPSSTENKAVSMTSVLSHSPGSGSTTGQGVTLAPTEPA 87
DB 1570 STPEAKAGVTSAPDXKRPXKAPXAGVTSAPDXKRPXKAPXAGVTSAPDXKRPX 1627
QY 88 SGGAATWQODVTVTPTRPALGSTTPRADVTSAPDXKRPXKAPXAGVTSAPDXKRPX 147
DB 1628 PGSTAPXAHGVTAPDXKRPXKAPXAGVTSAPDXKRPXKAPXAGVTSAPDXKRPX 1667
QY 148 PGSTAPXAHGVTAPDXKRPXKAPXAGVTSAPDXKRPXKAPXAGVTSAPDXKRPX 207
DB 1668 PGSTAPXAHGVTAPDXKRPXKAPXAGVTSAPDXKRPXKAPXAGVTSAPDXKRPX 1747
QY 208 SASIVNHGTSARATTPPAKSTPPSPSHSDPTTLAASHTKTDASSTHSTVPLPLS 267
DB 1748 SASIVNHGTSARATTPPAKSTPPSPSHSDPTTLAASHTKTDASSTHSTVPLPLS 1807
QY 268 SNHSTSPOLSTGVSPFFLSFPHISNLQFNLSLDPSTDYVLOLORDISBM 316
DB 1808 SNHSTSPOLSTGVSPFFLSFPHISNLQFNLSLDPSTDYVLOLORDISBM 1856

RESULT 7
```

```

US-08-783-774-2
Sequence 2; Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NON-SPLICING VARIANTS OF
TITLE OF INVENTION: GP350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
COMPUTER TYPE: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTERO Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-669-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 907 amino acids
TYPE: amino acid
STRANDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-783-774-2

Query Match 9.2%; Score 234.5; DB 3; Length 907;
Best Local Similarity 23.1%; Pred. No. 9.3e-10;
Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;

QY 22 TSGGHAASST-----PGGKERTATQSSVPSSTENKAVSMTSVLSHSPGSGSTT 73
DB 444 TTGTPSPSTHVPNTLTPASTGPTVST--ADVTSPTPAGTTSAGSPVSPSPMDNGTES 501
QY 74 QGQDVT--LAPTEPASGAATWQGVTSVP-----VTRPALGSTTEPPAH 116
DB 502 KAPMTSTSPVTTTPPNATSPPAVTTTPPNATSPPAVTTTPPNATSPPLGKTSPTSA 561
QY 117 DVTSAPD-NKPAFGSTAPPAHGVTSAPD--TRPPGSTAPPAHGVTSAPD--TRPAPGSTAP 173
DB 562 VTTTPPNATSPPLGKTSPTSAVTTTPPNATSPPLGKTSPTSAVTTTPPNATSPVIGSTP 621
QY 174 PARG-----VTSAPDNRPALGSTAPPAHGVTSAGSA-----SGSASTV--- 212
DB 622 QANNTNHTAGTSPTPVVTSQPNATSAVTTGQ--NHTSSSTSSNGLAPSSNHTLSP 679
QY 213 -VNHGTSARATTPA-----SKGTPSPSPSHSDPTTLAASHTKTDASSTHST 261
DB 680 TSDNSTHMPPLTSAHPTGENTGVTPPASISTHVTSSSPERPRTSQAGCGONST 739
QY 262 -----VPLPLSSNHSTSPOLSTGVSPFFLSFPHISNLQFNLSL-----EDP 301
DB 740 TKPEVAVVTKGTPPONATSPQASQKTAVPVTSQKANSITGKATGARTSTER 799
QY 302 STDYVLOLORDISEWFLQIVKQAGFLGSLNKKRPQSVVVOVLTIAFREQTINVAHDVETOR 361
DB 800 TTDY-----QGDSTTPRPRRYNATVLPST----- 824
```

QY 362 NOYTEASRYNLITSDVSVSH--VPPPSAQSGAGVPGW-GIALVLV-CVLVALAIY 417
 Db 825 ---SSKLRRFWTFTSPPTTAQATVPVPTSQ-----FRSNLSMLVLQMASLAVLTL 876
 QY 418 LIALAVCCRRKRYGQLDIFPARDTHPMSEYFTYTHGRVYPPSTDRSFY 469
 Db 877 LVMADCAFR-----NLSTSH---YTPPYDAEY 906

RESULT 8

US-09-328-599A-1
 Sequence 1, Application US/09328599A
 Patent No. 6432679
 GENERAL INFORMATION:
 APPLICANT: MOND, James J. and Lees, Andrew
 TITLE OF INVENTION: Enhancement of B Cell Activation by
 TITLE OF INVENTION: Co-ligation of Receptors for Antigen and Complement C3d
 TITLE OF INVENTION: Using EBV gp350/220 or EBV gp350/220 Peptide Adjuvants
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 ADDRESSEE: Dunnet, L.L.P.
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/328,599A
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Forde, Jean B.
 REGISTRATION NUMBER: 32,984
 REFERENCE/DOCKET NUMBER: 04995, 6025-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)408-4000
 TELEFAX: (202)408-4000
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 907 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-328-599A-1

Query Match 9.2%; Score 234.5; DB 4; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9.3e-10;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;

QY 22 TSGSHAST-----PGKEKTSATORSVPSSTKNAVSMTSSVLSHSPSGSSST 73
 Db 444 TTGTPSSTHVPTNLTPASTGPTVST--ADVTSPTAGTSGAPVTPSPMDNGTES 501
 QY 74 QGQDVT--LAPATEPASGAATWGDVTVP-----VTRPALGSTTPAH 116
 Db 502 KAPDMISSTSVTTPPNKTSPTPAVTPPTPNATSPPTATTPPNATSTPLKTSPTSA 561
 QY 117 DVTSPAD-NKPAGSTAPAGVTSAPD--TRPPGSTAPAPAGVTSAPD--TRAPGSTAP 173
 Db 562 VTTPPNATSPPLTGKTSPTSAVTPPTPNATSPPLTGKTSPTSAVTPPTPNATGPTVQETS 621
 QY 174 PANG-----VTSAPDRRPLAGTAPVNAHVTSASGSA-----SGSASTL-- 212
 Db 622 QANATNHTLAGTSPTPVTSQPKNATSAVTGO--HNTSSSTSSMSLRPSNPTLSPS 679

QY 213 -VANGTSARATTTA-----SKTSPSISHSDDPTTLASHTKDSSTHST-- 261
 Db 680 TSNNGTSHMPLTISAPPTGNGINIOVTPASISTHVTSSBPRRGTTSSQSGKSSN 739
 QY 262 ---VPLTSSNISTSPOLSTGVSPFPLSFHLSNLOPNSL-----EDP 301
 Db 740 TRGEVNVTKGTTPQVATSPQAPSGQKTAIVPTVSTGKANSYTGKATTHGARTSP 799
 QY 302 STDYQELQRDISEMEFLQIKXGQFGLSNIKFRPGSVVQLTIAFRREGTINADVETOF 361
 Db 800 TTDY-----GDSTTPPRRYNATVLPST----- 824
 QY 362 NOYTEASRYNLITSDVSVSH--VPPPSAQSGAGVPGW-GIALVLV-CVLVALAIY 417
 Db 825 ---SSKLRRFWTFTSPPTTAQATVPVPTSQ-----FRSNLSMLVLQMASLAVLTL 876
 QY 418 LIALAVCCRRKRYGQLDIFPARDTHPMSEYFTYTHGRVYPPSTDRSFY 469
 Db 877 LVMADCAFR-----NLSTSH---YTPPYDAEY 906

RESULT 9

PCT-US95-04611A-19
 Sequence 19, Application PC/TUS9504611A
 GENERAL INFORMATION:
 APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
 TITLE OF INVENTION: Non Splicing Variants of gp350/220
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 STREET: 5 Palo Alto Square
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04611A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/229,291
 FILING DATE: April 18, 1994

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Luann Ceert
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: AVIR-003/00US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-843-5163
 TELEFAX: 415-857-0663
 TELEX: 380816 CooleyPA
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 907 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-04611A-19

Query Match 9.2%; Score 234.5; DB 5; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9.3e-10;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;

QY 22 TSGSHAST-----PGKEKTSATORSVPSSTKNAVSMTSSVLSHSPSGSSST 73
 Db 444 TTGTPSSTHVPTNLTPASTGPTVST--ADVTSPTAGTSGAPVTPSPMDNGTES 501
 QY 74 QGQDVT--LAPATEPASGAATWGDVTVP-----VTRPALGSTTPAH 116

```

Db 502 KAFDSTNHSIPVTTTPNATSPPTPAVTTTPNANISPTPAVTTTPNATSPPLGTGKTSSTA 561
QY 117 DVTSAPD-KKPAFGSTAPRAHQMSAPD--TRPPGSTAPAAHGVSAPD-TRAPASAP 173
Db 562 VTTTPNATSPPLGTGKTSPTSAVTTTPNATSPPLGTGKTSPTSAVTTTPNATSPPLGTGTS 621
QY 174 PPHG-----VTSADNRPLAGSTAPPHVHVSASGSA-----SGSASTL--- 212
Db 622 QANATNHTLGTGSPVVTISQPKATSAVTTGCO--HNITSSSTSHMELARSNSPPLTSPS 679
QY 213 -VHAGTSASATTTA-----SKTSPSPISNHSSTPTTLASHTKIDASHTHST-- 261
Db 680 TSDNSTSHMPLTTSNHPGTGENTIOVTPASISTHNVSTSPSPRPCTTSGASGNSSTGS 739
QY 262 -----VPLTSSNHSSTPGLSTGVSFFPLTSFHLNLOQNSSTL-----EDP 301
Db 740 TRGEVAVTKGTIPNNATSIQONPSQOKTAPATVPTSGKANSITGGKHTTGARTSTEP 799
QY 302 SPDYVQGLQDISNEMFOIYNGGCEGLASNIKFRSGSVVOLTAFRBGTINHDVETOP 361
Db 800 TTDV-----GDSYTRPRRYNATYLPST----- 824
QY 362 NQYTEMASRYVLTISGVSVSH--VPPPSASGAGVPCW-GIALVIV-CVLAVALIYV 417
Db 825 ---SSKTRPRRTFTSPVTTAAQATVPVPTSG---PFRSMLMVLQMSALVLTLL 876
QY 418 LIALAVOCCRRKNYGOLDIFPRADTYPHMSSEFYVTHGRVYPPSSITDSP 469
Db 877 LTVAMDCAFR-----NLSTSHT-----VTPPEYDDARY 906

RESULT 10
US-09-556-706B-2
Sequence 2, Application US/09556706B
General Information:
Applicant: Spaele, Richard
Applicant: Jackman, Minthrop
Title of Invention: NON SLICING VARIANTS OF GP350/220
File Reference: 7682-050-999
Current Application Number: US/09/556, 706B
Current Filing Date: 2006-04-24
Prior Application Number: 08/783, 774
Prior Filing Date: 1997-01-15
Prior Application Number: 08/229, 291
Prior Filing Date: 1994-04-18
Number of Seq ID NOS: 19
Software: Patent version 3.0
Seq ID NO 2
Length: 878
Type: PRT
Organism: Virus
Other Information: gp350
US-09-556-706B-2

Query Match 8.9%; Score 228; DB 4; Length 878;
Best Local Similarity 22.8%; Pred. No. 2.8e-09;
Matches 118; Conservative 60; Mismatches 187; Indels 152; Gaps 211

QY 22 TGSQHASST-----PCGEKETASQSSVSPSSTKNAVSMTSSVLSNPSQSSSTT 73
Db 444 TTTGLDSHPVTLTAIPASTGPVSI--ADVTSPFAGTSGASVPTPSPSPMDNGTGS 501
QY 74 QGQDVT--LAPATEPASGSAMTQGDVTSVP-----VTRPALGSTTPPAH 116
Db 502 KAPDMSTSTSPVTTTPNATSPPTPAVTTTPNATSPPTPAVTTTPNATSPPLGTGTSSTA 561
QY 117 DVTSAPD-KKPAFGSTAPRAHGVSAPD--TRPPGSTAPAAHGVSAPD-TRAPASAP 173
Db 562 VTTTPNATSPPLGTGKTSPTSAVTTTPNATSPPLGTGKTSPTSAVTTTPNATSPPLGTGTS 621
QY 174 PPHG-----VTSADNRPLAGSTAPPHVHVSASGSA-----SGSASTL--- 212
Db 622 QANATNHTLGTGSPVVTISQPKATSAVTTGCO--HNITSSSTSHMELARSNSPPLTSPS 679
QY 213 -VHAGTSASATTTA-----SKTSPSPISNHSSTPTTLASHTKIDASHTHST-- 261
Db 680 TSDNSTSHMPLTTSNHPGTGENTIOVTPASISTHNVSTSPSPRPCTTSGASGNSSTGS 739
QY 262 -----VPLTSSNHSSTPGLSTGVSFFPLTSFHLNLOQNSSTL-----EDP 301
Db 740 TRGEVAVTKGTIPNNATSIQONPSQOKTAPATVPTSGKANSITGGKHTTGARTSTEP 799
QY 302 SPDYVQGLQDISNEMFOIYNGGCEGLASNIKFRSGSVVOLTAFRBGTINHDVETOP 361
Db 800 TTDV-----GDSYTRPRRYNATYLPST----- 824
QY 362 NQYTEMASRYVLTISGVSVSH--VPPPSASGAGVPCW-GIALVIV-CVLAVALIYV 417
Db 825 ---SSKTRPRRTFTSPVTTAAQATVPVPTSG---PFRSMLMVLQMSALVLTLL 876
QY 418 LIALAVOCCRRKNYGOLDIFPRADTYPHMSSEFYVTHGRVYPPSSITDSP 469
Db 877 LTVAMDCAFR-----NLSTSHT-----VTPPEYDDARY 906

```

| QY | 174 | PAHGVTSADNRKPA | LSGTSF-- | RYNATYASAGSAGSASGLVHNTSARATTPYSSK-- | 229 |
|----|-----|---|--|-------------------------------------|---------|
| Db | 622 | QANATKN----- | TLGTSFTVTV-- | VTSPQKNAISATVYQGNRPNSSNPETLISPTSDNS | 673 |
| QY | 230 | ----- | TPPSISGHSDIPTPLTSLASTYTDASSTHST-- | ----- | VPPLTSS |
| Db | 674 | TSNHGNGNITVQTPALSTHIVSTHSSTSEPRRCOTSSQSPONSSTIRKGEVNTKTPR | 733 | | |
| QY | 269 | NHSTSPDLSGTGSEFFLSPHLSNCRNSSTL-- | SDPTTYVQELQRISE | 315 | |
| Db | 734 | QNAISPPASGCKTAVPTVTSYGGKANSTYTGKTHGACRTSTPTDY-- | ----- | 783 | |
| QY | 316 | MFGLVYKGGFGLSNIKRFGSVVQULLAFREGTINHVDFOTFQNYKTEAASRYNL | 375 | | |
| Db | 784 | ----- | CGSITPRPRRYNTLTLPSTP-- | SSGLPWRFT | 815 |
| QY | 376 | ISDVSVH-- | VEPPPSAGSGAGVPGW-GIALVLCVLAIVYITALAVCCCRKNYG | 432 | |
| Db | 816 | SPPTTKQATVPEFTSQ----- | FRSNLSMTLV----- | LTLLVNAIDCARH-- | 858 |
| QY | 433 | QLDIFPARQYTHMSEVPTVYTHORVYVSPSTDSXY | 469 | | |
| Db | 859 | ----- | NLSHST-- | YTPPYDALEY | 877 |

| QY | 174 | PAHGVTSADNRKPA | LSGTSF-- | RYNATYASAGSAGSASGLVHNTSARATTPYSSK-- | 229 |
|----|-----|---|--|-------------------------------------|---------|
| Db | 622 | QANATKN----- | TLGTSFTVTV-- | VTSPQKNAISATVYQGNRPNSSNPETLISPTSDNS | 673 |
| QY | 230 | ----- | TPPSISGHSDIPTPLTSLASTYTDASSTHST-- | ----- | VPPLTSS |
| Db | 674 | TSNHGNGNITVQTPALSTHIVSTHSSTSEPRRCOTSSQSPONSSTIRKGEVNTKTPR | 733 | | |
| QY | 269 | NHSTSPDLSGTGSEFFLSPHLSNCRNSSTL-- | SDPTTYVQELQRISE | 315 | |
| Db | 734 | QNAISPPASGCKTAVPTVTSYGGKANSTYTGKTHGACRTSTPTDY-- | ----- | 783 | |
| QY | 316 | MFGLVYKGGFGLSNIKRFGSVVQULLAFREGTINHVDFOTFQNYKTEAASRYNL | 375 | | |
| Db | 784 | ----- | CGSITPRPRRYNTLTLPSTP-- | SSGLPWRFT | 815 |
| QY | 376 | ISDVSVH-- | VEPPPSAGSGAGVPGW-GIALVLCVLAIVYITALAVCCCRKNYG | 432 | |
| Db | 816 | SPPTTKQATVPEFTSQ----- | FRSNLSMTLV----- | LTLLVNAIDCARH-- | 858 |
| QY | 433 | QLDIFPARQYTHMSEVPTVYTHORVYVSPSTDSXY | 469 | | |
| Db | 859 | ----- | NLSHST-- | YTPPYDALEY | 877 |

| QY | 174 | PAHGVTSADNRKPA | LSGTSF-- | RYNATYASAGSAGSASGLVHNTSARATTPYSSK-- | 229 |
|----|-----|---|--|-------------------------------------|---------|
| Db | 622 | QANATKN----- | TLGTSFTVTV-- | VTSPQKNAISATVYQGNRPNSSNPETLISPTSDNS | 673 |
| QY | 230 | ----- | TPPSISGHSDIPTPLTSLASTYTDASSTHST-- | ----- | VPPLTSS |
| Db | 674 | TSNHGNGNITVQTPALSTHIVSTHSSTSEPRRCOTSSQSPONSSTIRKGEVNTKTPR | 733 | | |
| QY | 269 | NHSTSPDLSGTGSEFFLSPHLSNCRNSSTL-- | SDPTTYVQELQRISE | 315 | |
| Db | 734 | QNAISPPASGCKTAVPTVTSYGGKANSTYTGKTHGACRTSTPTDY-- | ----- | 783 | |
| QY | 316 | MFGLVYKGGFGLSNIKRFGSVVQULLAFREGTINHVDFOTFQNYKTEAASRYNL | 375 | | |
| Db | 784 | ----- | CGSITPRPRRYNTLTLPSTP-- | SSGLPWRFT | 815 |
| QY | 376 | ISDVSVH-- | VEPPPSAGSGAGVPGW-GIALVLCVLAIVYITALAVCCCRKNYG | 432 | |
| Db | 816 | SPPTTKQATVPEFTSQ----- | FRSNLSMTLV----- | LTLLVNAIDCARH-- | 858 |
| QY | 433 | QLDIFPARQYTHMSEVPTVYTHORVYVSPSTDSXY | 469 | | |
| Db | 859 | ----- | NLSHST-- | YTPPYDALEY | 877 |

| QY | 174 | PAHGVTSADNRKPA | LSGTSF-- | RYNATYASAGSAGSASGLVHNTSARATTPYSSK-- | 229 |
|----|-----|---|--|-------------------------------------|---------|
| Db | 622 | QANATKN----- | TLGTSFTVTV-- | VTSPQKNAISATVYQGNRPNSSNPETLISPTSDNS | 673 |
| QY | 230 | ----- | TPPSISGHSDIPTPLTSLASTYTDASSTHST-- | ----- | VPPLTSS |
| Db | 674 | TSNHGNGNITVQTPALSTHIVSTHSSTSEPRRCOTSSQSPONSSTIRKGEVNTKTPR | 733 | | |
| QY | 269 | NHSTSPDLSGTGSEFFLSPHLSNCRNSSTL-- | SDPTTYVQELQRISE | 315 | |
| Db | 734 | QNAISPPASGCKTAVPTVTSYGGKANSTYTGKTHGACRTSTPTDY-- | ----- | 783 | |
| QY | 316 | MFGLVYKGGFGLSNIKRFGSVVQULLAFREGTINHVDFOTFQNYKTEAASRYNL | 375 | | |
| Db | 784 | ----- | CGSITPRPRRYNTLTLPSTP-- | SSGLPWRFT | 815 |
| QY | 376 | ISDVSVH-- | VEPPPSAGSGAGVPGW-GIALVLCVLAIVYITALAVCCCRKNYG | 432 | |
| Db | 816 | SPPTTKQATVPEFTSQ----- | FRSNLSMTLV----- | LTLLVNAIDCARH-- | 858 |
| QY | 433 | QLDIFPARQYTHMSEVPTVYTHORVYVSPSTDSXY | 469 | | |
| Db | 859 | ----- | NLSHST-- | YTPPYDALEY | 877 |

| QY | 174 | PAHGVTSADNRKPA | LSGTSF-- | RYNATYASAGSAGSASGLVHNTSARATTPYSSK-- | 229 |
|----|-----|---|--|-------------------------------------|---------|
| Db | 622 | QANATKN----- | TLGTSFTVTV-- | VTSPQKNAISATVYQGNRPNSSNPETLISPTSDNS | 673 |
| QY | 230 | ----- | TPPSISGHSDIPTPLTSLASTYTDASSTHST-- | ----- | VPPLTSS |
| Db | 674 | TSNHGNGNITVQTPALSTHIVSTHSSTSEPRRCOTSSQSPONSSTIRKGEVNTKTPR | 733 | | |
| QY | 269 | NHSTSPDLSGTGSEFFLSPHLSNCRNSSTL-- | SDPTTYVQELQRISE | 315 | |
| Db | 734 | QNAISPPASGCKTAVPTVTSYGGKANSTYTGKTHGACRTSTPTDY-- | ----- | 783 | |
| QY | 316 | MFGLVYKGGFGLSNIKRFGSVVQULLAFREGTINHVDFOTFQNYKTEAASRYNL | 375 | | |
| Db | 784 | ----- | CGSITPRPRRYNTLTLPSTP-- | SSGLPWRFT | 815 |
| QY | 376 | ISDVSVH-- | VEPPPSAGSGAGVPGW-GIALVLCVLAIVYITALAVCCCRKNYG | 432 | |
| Db | 816 | SPPTTKQATVPEFTSQ----- | FRSN | | |

Thu May 8 16:14:00 2003

us-09-658-621b-2.ra1

Page 9

Db 531 VTTGVVVISEDVTTCPLTETTPATESAPATESAPATESAPATESAPTES 590
Qy 246 ASHSTKTDASTHSTVPLTSSN---HSTSP 274
Db 591 TAPATES-APATESVPACGTSSVAPESAP 622

RESULT 12

US-08-099-354-1
Sequence 1, Application US/08099354
Patent No. 574144
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELLARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/099,354
FILING DATE: 30-JUL-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STRILLA, GEORGE M.
REGISTRATION NUMBER: 18221
REFERENCE/DOCKET NUMBER: 6137/202246
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3536
TELEFAX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-099-354-1

Query Match 8.0%; Score 205; DB 1; Length 40;
Best Local Similarity 95.0%; Pred. No. 2.1e-09;
Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 141 APDTPPGSTAPAPAGVTSAPDTPAPAGSTAPAPAGVTS 180
Db 1 APDTPAPAGSTAPAPAGVTSAPDTPAPAGSTAPAPAGVTS 40

RESULT 13

US-08-288-059-7
Sequence 7, Application US/08288059
Patent No. 582766
GENERAL INFORMATION:
APPLICANT: FINN, OLIVERA J.
APPLICANT: FONTENOT, J. D.
APPLICANT: MONTELLARO, RONALD C.
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/288,059
FILING DATE: 08-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARILAN K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 61137/205204
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3711
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-288-059-7

Query Match 8.0%; Score 205; DB 2; Length 40;
Best Local Similarity 95.0%; Pred. No. 2.1e-09;
Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 141 APDTPPGSTAPAPAGVTSAPDTPAPAGSTAPAPAGVTS 180
Db 1 APDTPAPAGSTAPAPAGVTSAPDTPAPAGSTAPAPAGVTS 40

RESULT 14
US-08-911-393-2
Sequence 2, Application US/08911393
Patent No. 6323008
GENERAL INFORMATION:
APPLICANT: PELLETIER, Marc
APPLICANT: BARKER, William A.
APPLICANT: HAKES, David J.
APPLICANT: ZOPF, David A.
TITLE OF INVENTION: METHODS FOR PRODUCING
TITLE OF INVENTION: STABILIZED SACCHARIDES IN A DAIRY SOURCE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,393
FILING DATE: 14-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-032-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8669741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-911-393-2

Query Match 8.0%; Score 205; DB 4; Length 1060;
Best Local Similarity 27.9%; Pred. No. 2, 1e-07;
Matches 88; Conservative 31; Mismatches 136; Indels 60; Gaps 16;

QY 16 TVTITVTSAGHASTPGGEKETSATQSSVPESTKNAVSTSVLSHSPGSGSTTQG 75
DB 688 TPSTPADSSAH--STPSTPADSSAHSTPSAPDNGAHSTPSTPGDSSAHSTPSTADNGA 745
QY 76 QDVTLAPATEPAGSAAAT---WGQDVTSPVTRPALGSTTPP---AHDVTSAPDNKPA 127
DB 746 HSTPSAPADSNHSTPSTPADNGAHSTPSTPADNGAHSTPSTPADNGAHSTPSTPGDSSA 805
QY 128 PGSTAPP---AHGVTSAP-----DTRPPPGS-----TAPR---AHGVTSAP----- 162
DB 806 HSTPSTPADNGAHSTPSAPADSNHSTPSTPGDNGAHSTPSAPADSNHSTPSTADSSA 865
QY 163 -DTRPAPGS-----TRP---PAHGVTSAPDNKRALGSTRPP---VANNVTSAGSASG 207
DB 866 HSTPSAPAGNGAHSTPSAPADSNHSTPSAPDNGAHSTPSAPADNGAHSTPSAPADSSA 925
QY 208 SASTLVHNGTSARAT--TTPASKS--TPSPISPH--HS--DTPTLASHSTKT--DASST 257
DB 926 HSTPSTPADSSAHSTPSTPADSSAHSTPSAPDNGAHSTPSAPADSSAHSTPSTPGDSSA 985
QY 258 HHSTVPEPLTSSNHST 272
DB 986 HSTPSAPADSSAHST 1000

RESULT 15

US-09-134-001C-4463
Sequence 4463 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4463
LENGTH: 2137
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 7.7%; Score 198; DB 4; Length 2137;
Best Local Similarity 23.9%; Pred. No. 1, 9e-06;
Matches 108; Conservative 71; Mismatches 181; Indels 92; Gaps 15;

QY 19 TVTTSAGHASTPGGEKETSATQSSVPESTKNAVSTSVLSHSPG----- 67

DB 1699 TSLSDSTSTSESGSTSTSESDSDASSTSLSESTSTSLSDSTSTSTSDASSTSMVSDS 1758
QY 68 -----SGSSTTQGVDTLAPATEP-----SGSAATWQDVTSPVTRPALGSTT 112
DB 1759 NRASTSLSDSTSVSDSTASSTSEASSTSESTSTASSTSLSESTSVSDSTSTST 1818
QY 113 PPAHDVTSAPDNKPAAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPA----P 168
DB 1819 DSASTSTSESDNSR--STSLSESTSVSDS--TSASTSASASTSVSDSNASASTSL 1874
QY 169 GSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAGSAGSASTLVHNGTS-----ARATT 223
DB 1875 GSTSTSVSDSTST--STASASTSTSESDSDASTSLSGSTSTSLSDSTSTSDASST 1932
QY 224 TPAKSTPPSPISPHSDPTTLASHSTKTDASTHSTVPEPLTSSNHSTSPQLSTGVSFF 283
DB 1933 TSEASTSTSESDSESDSESTSVSESDSTSVSDSTSTSTSEASASTSTSESTSESTSV 1990
QY 284 FLSFHISNLQFNSSLDPSTDYQELORDISEMFLQ---LYKQGFILGLSNIKFRPGSV 340
DB 1991 -----EBSSTSLSDSS--TSTSMSTSEFTPSQSPINSSEQFIDGSD-----L 2030
QY 341 VQTLAEREGTINV-----HDVEYQFNQYKTEASRYNLTISDVSVSH----- 383
DB 2031 SEDTIVTQSKNTMTMLNKTGKDYDLOEQRGYTDS---QHNETOQMDNHSNNLDLHQ 2087
QY 384 -----VPPPSAQSAGAVPGMGIALLVLCV 409
DB 2088 RLQDKVVKQPTGDEGVVSNGLVAVLVLA 2119

Search completed: May 1, 2003, 23:14:26
Job time : 34.5357 secs

07 341 пртптт лсуствтлн естпчествуды теснпствудой стгуяеуеет сеутент оенсст ед 300

| | | | |
|----|-----|--|-----|
| D6 | 241 | TTTTAASTTTDAASTHSTATPPLTSSNNSTBPOLSTGVSFPFLASHISNLOFNNSLTD | 300 |
| QY | 301 | PSTDYVYOLORDISEMFLIYVQGGFGLANIFRRGSVVQLTLAREBGTINVDVETQ | 360 |
| D6 | 301 | PSTDYVYOLORDISEMFLIYVQGGFGLANIFRRGSVVQLTLAREBGTINVDVETQ | 360 |
| QY | 361 | FNQYTYEASRYNLTISDVSVSHVPFPESAQAGVPGMGIALVYLVCVYVLAATVYLA | 420 |
| D6 | 361 | FNQYTYEASRYNLTISDVSVSHVPFPESAQAGVPGMGIALVYLVCVYVLAATVYLA | 420 |
| QY | 421 | LAVCCCRKNKNGQLTFPPARDVTHPMSRYPTVTHGRVYVPSSTDRSPYKESACAGGSS | 480 |
| D6 | 421 | LAVCCCRKNKNGQLTFPPARDVTHPMSRYPTVTHGRVYVPSSTDRSPYKESACAGGSS | 480 |
| QY | 481 | LSTYTPPAVAATSAATL | 495 |
| D6 | 481 | LSTYTPPAVAATSAATL | 495 |

```

RESULT 2
US-09-658-621B-67
: Sequence 67, Application US/09658621B
: GENERAL INFORMATION:
: APPLICANT: Taylor-Papadimitriou, Joyce
: APPLICANT: Heukamp, Lukas Carl
: APPLICANT: Offringa, Rienk
: APPLICANT: Melief, Cornelis Johanna Maria
: APPLICANT: Acres, Bruce
: APPLICANT: Thomas, Michelle
: TITLE OF INVENTION: MHC-1 derived peptides
: FILE REFERENCE: 029395-017
: CURRENT FILING DATE: 2000-09-08
: PRIOR APPLICATION NUMBER: US/09/658, 621B
: PRIOR FILING DATE: 2000-03-03
: PRIOR APPLICATION NUMBER: GB 9922142.5
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: EP 99 40 2237.4
: PRIOR FILING DATE: 1999-09-10
: NUMBER OF SEQ ID NOS: 80
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO: 67
: LENGTH: 475
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-658-621B-67

```

| Query Match | 94.2%; | Score 21.3; | DB 5; | Length 475; |
|-----------------------|--------------|--|---------------|-------------|
| Best Local Similarity | 95.6%; | Pred. No. 5.7e-97; | | |
| Matches 473; | Conservative | 0; | Mismatches 2; | Indels 20; |
| | | | | Gaps |
| Qy | 1 | MPFGQSPFFLLLLTLVLTVTGSGHASTPGGEXETASATQSSVPSSTENAMVSTSV | 60 | |
| Dp | 1 | MPFGQSPFFLLLLTLVLTVTGSGHASTPGGEXETASATQSSVPSSTENAMVSTSV | 60 | |
| Qy | 61 | LSHSPGSGSSTTGQDYLPAERPPSSAAWVGQDVTSPVTPPALGSTPPADDT | 120 | |
| Dp | 61 | LSHSPGSGSSTTGQDYLPAERPPSSAAWVGQDVTSPVTPPALGSTPPADDT | 120 | |
| Qy | 121 | ADPNKPAGSTAPRAGTSAVDTRPPGSTAPRAHVSTAPDTPAAGSTAPRAGTS | 180 | |
| Dp | 121 | ADPNKPAGSTAPRAGTSAVDTRPPGSTAPRAHVSTAPDTPAAGSTAPRAGTS | 180 | |
| Qy | 181 | ADPNPALGSTAPRPNYTSAGSAGSAGSASTLYNNGSAAATTPAGSKTPBET | 240 | |
| Dp | 161 | ADPNPALGSTAPRPNYTSAGSAGSAGSASTLYNNGSAAATTPAGSKTPBET | 220 | |
| Qy | 241 | TPPTLAHSIKTKTAASTHSTVPPPLTSSNMHSTPQSTGVSGPFFLSPTISLQPNLSLD | 300 | |
| Dp | 221 | TPPTLAHSIKTKTAASTHSTVPPPLTSSNMHSTPQSTGVSGPFFLSPTISLQPNLSLD | 280 | |
| Qy | 301 | PSTDIYQELQDISEMFLQIYKQGGFLGLSNIKRPPSVVQLTLAIREGTINVDVETQ | 360 | |

| | | | |
|----|-----|---|-----|
| Db | 281 | PSTDYPPLOHRIIDISEMFQIQIYKGGPFGISNIKFFPGGSVVUULLAFREGTINVHDETQ | 340 |
| Qy | 361 | FMQYETEAASNYLITISDVSYASHPPFPFSASQDAGVPMQIALIVYVCLVALVALAVLYIA | 420 |
| Db | 341 | FMQYETEAASNYLITISDVSYSDVPPFPFSASQDAGVPMQIALIVYVCLVALVALAVLYIA | 400 |
| Qy | 421 | LAVCCCRKKNYGOLDIFPARDTYPMSEEPYTHYHGRVYVPPSSYDTRSPEYKYSAGNGSS | 480 |
| Db | 401 | LAVCCCRKKNYGOLDIFPARDTYPMSEEPYTHYHGRVYVPPSSYDTRSPEYKYSAGNGSS | 460 |
| Qy | 481 | LSYTNPAVAATSNL | 495 |
| Db | 461 | LSYTNPAVAATSNL | 475 |

```

RESULT 3
US-10-348-119-309
; Sequence 309, Application US/10348119
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASES
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 NP
; CURRENT APPLICATION NUMBER: US/10/348,119
; CURRENT FILING DATE: 2003-01-17
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 309
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-348-119-309

```

| Query Match | 94.2% | Score 2413 | DB 6 | Length 475 |
|-----------------------|--------------|--|--------------|------------|
| Best Local Similarity | 95.6% | Pred. No 5,7e-97 | | |
| Matches 473 | Conservative | 0 | Mismatches 2 | Indels 20 |
| Gaps | | | | |
| Qy | 1 | MPGQSPFLILLLTLVTVTGSGASSTPGKEKESATQKRSVSPSTSKNAVMTSSV | 60 | |
| Db | 1 | MPGQSPFLILLLTLVTVTGSGASSTPGKEKESATQKRSVSPSTSKNAVMTSSV | 60 | |
| Qy | 61 | LSHSPPSSSSTTQGDVTLAPTPBPAGSAAITGQDVTSVVTPRALPGSTTPPAHDVTS | 120 | |
| Db | 61 | LSHSPPSSSSTTQGDVTLAPTPBPAGSAAITGQDVTSVVTPRALPGSTTPPAHDVTS | 120 | |
| Qy | 121 | ADPNKPAAGSTAPPAHGVTSAPDTPRPGSTAPPAHGVTSAPDTPRALPGSTTPPAHGVTS | 180 | |
| Db | 121 | ADPNKPAAGSTAPPAHGVTSAPDTPRPGSTAPPAHGVTSAPDTPRALPGSTTPPAHGVTS | 160 | |
| Qy | 181 | ADPNRPAIGSTAPPAHNTVTSAGSAGSAGSITLVNNGSARAATTTPAKSPPTGISHSND | 240 | |
| Db | 161 | ADPNRPAIGSTAPPAHNTVTSAGSAGSAGSITLVNNGSARAATTTPAKSPPTGISHSND | 220 | |
| Qy | 241 | PTTTLASHKTKDASSTHSVSPRLTSSMSHSTSGCTSGVSEFLSHSISNLOPSSSED | 300 | |
| Db | 221 | PTTTLASHKTKDASSTHSVSPRLTSSMSHSTSGCTSGVSEFLSHSISNLOPSSSED | 280 | |
| Qy | 301 | PESTDVYQELORDISEMFLQITKYCGGFLGLSNIKRPGSSVYVUULIARREGTINVDYEQ | 360 | |
| Db | 281 | PESTDVYQELORDISEMFLQITKYCGGFLGLSNIKRPGSSVYVUULIARREGTINVDYEQ | 340 | |
| Qy | 361 | PNQYKTPASRNLTITSVSYSVHPPPPSAOSGAGVPGMGIALVLVCVLVALAIVYLIA | 420 | |
| Db | 341 | PNQYKTPASRNLTITSVSYSVHPPPPSAOSGAGVPGMGIALVLVCVLVALAIVYLIA | 400 | |
| Qy | 421 | LAVCCRRKRVGQDLIPFADTYHPMSFVYTHHGRVYPPSTDRSPYKVSAGNCGSS | 480 | |
| Db | 401 | LAVCCRRKRVGQDLIPFADTYHPMSFVYTHHGRVYPPSTDRSPYKVSAGNCGSS | 460 | |

Thu May, 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 3

```
QY      481  LSYTNPAVVAATSA NL  495
         |||||
Db      461  LSYTNPAVVAATSA NL  475
```

RESULT 4
US-60-42

? Sequence 59, Application US/60422176
 ? GENERAL INFORMATION:
 ? APPLICANT: Pharmacia Corporation
 ? APPLICANT: Boumter, Maureen J.
 ? TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES INVOLVED IN CANCER. THE POLYPEPTIDE
 ? TITLE OF INVENTION: ENCODED THERAPY, AND METHODS OF USING THE SAME
 ? FILE REFERENCE: 01040 PR
 ? CURRENT APPLICATION NUMBER: US/60/422,176
 ? CURRENT FILING DATE: 2002-10-29
 ? NUMBER OF SEQ ID NOS: 314
 ? SOFTWARE: PatentIn version 3.1
 ? SEQ ID NO 59
 ? LENGTH: 175
 ? TYPE: CDS
 ? ORGANISM: homo sapiens
 ? US-60-422176-59

| | | | | |
|---------------------------|--------|--------------------|------------|-------------|
| Query Match | 94.2%; | Score 2413; | DB 7; | Length 475; |
| Best Local Similarity | 95.6%; | Pred. No. 5.7e-97; | | |
| Matches 473; Conservative | 0; | Mismatches 2; | Indels 20; | Gaps 1 |

I

| | | | |
|----|-----|---|-----|
| Qy | 1 | MPBGOSFFLLLLVTVLVVNGSSASHPGGEKKSATYRSSVBSSTKAAVSMTSV | 60 |
| Db | 1 | MPBGOSFFLLLLVTVLVVNGSSASHPGGEKKSATYRSSVBSSTKAAVSMTSV | 60 |
| Qy | 61 | LSHSPGSGSSTTGGDVTTLATATEPAGSAAATWGQDVTSVPTVRPALGSTTPPAHDVS | 120 |
| Db | 61 | LSHSPGSGSSTTGGDVTTLATATEPAGSAAATWGQDVTSVPTVRPALGSTTPPAHDVS | 120 |
| Qy | 121 | APDNPPAGSTAPPAHVTSAPDTRPPGSTAAPAHAGVTSAPDTRPPAGSTAPPAHVTS | 180 |
| Db | 121 | APDNPPAGSTAPPAHVTSAPDTRPPGSTAAPAHAGVTSAPDTRPPAGSTAPPAHVTS | 180 |
| Qy | 181 | APDNPPALGSTAPPAHVNTVSASGSASGASATLVHNQTSARATTPPAKSPSPSISHSND | 240 |
| Db | 161 | APDNPPALGSTAPPAHVNTVSASGSASGASATLVHNQTSARATTPPAKSPSPSISHSND | 220 |
| Qy | 241 | TPPTLASHSTKTDASSTHSTVPPLTSSNHSSTPOLSTVSPFFLSHISINQFMSLED | 300 |
| Db | 221 | TPPTLASHSTKTDASSTHSTVPPLTSSNHSSTPOLSTVSPFFLSHISINQFMSLED | 280 |
| Qy | 301 | PSRDVYQLOPDISEMFLOIYQGGFGLNINIFPGSVVQTLARBSSTINHDVETQ | 360 |
| Db | 281 | PSRDVYQLOPDISEMFLOIYQGGFGLNINIFPGSVVQTLARBSSTINHDVETQ | 340 |
| Qy | 361 | FNQYKTAASRYNLTSPDVSHPFPFSAOSAGVPGMGALNLYMCTVALALIVYLA | 420 |
| Db | 341 | FNQYKTAASRYNLTSPDVSHPFPFSAOSAGVPGMGALNLYMCTVALALIVYLA | 400 |
| Qy | 421 | LAVCCRRKATGOLDIPPAORTHHMSYPRYTHKRYTPPSSTDSBPYKXISAENGSSS | 480 |
| Db | 401 | LAVCCRRKATGOLDIPPAORTHHMSYPRYTHKRYTPPSSTDSBPYKXISAENGSSS | 460 |
| Qy | 481 | LSTTPPAVAATSANL 495 | |
| Db | 461 | LSTTPPAVAATSANL 475 | |

RESULT 5
US-60-452-680-12835
; Sequence 12835, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILE, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

```

? FILE REFERENCE: CLO01450
? CURRENT APPLICATION NUMBER: US/60/452,688
? CURRENT FILING DATE: 2003-03-07
? NUMBER OF SEQ ID NOS: 16213
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 12833
? LENGTH: 475
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-60-452-680-12833

```

| | | | | |
|-----------------------|--------------|--------------------|---------------|-------------|
| Query March | 94.2%; | Score 2413; | DB 7; | Length 475; |
| Best Local Similarity | 95.6%; | Pred. No. 5.7e-97; | | |
| Matches 473; | Conservative | 0; | Mismatches 2; | Indels 20; |
| | | | | Gaps 1. |

| | | | |
|----|-----|--|-----|
| Qy | 1 | MTPGTGPFFLLILLTLVTLVTVTGSGHASSPGGEKETSATORSVBSSTEKUAVMTSSV | 60 |
| Db | 1 | MTPGTGPFFLLILLTLVTLVTVTGSGHASSPGGEKETSATORSVBSSTEKUAVMTSSV | 60 |
| Qy | 61 | LSHSPSSGSGSTTGGODVTLTPATEPAGSAITWQODTVSVVTRPALSGSTTTPPAHDVTS | 120 |
| Db | 61 | LSHSPSSGSGSTTGGODVTLTPATEPAGSAITWQODTVSVVTRPALSGSTTTPPAHDVTS | 120 |
| Qy | 121 | APDNKKPAPGSTATPAHGVTSAPDTRPPGCGTAPAAHGVTSAPDTRPPAGSTAPPAHGVTS | 180 |
| Db | 121 | APDNKKPAPGSTATP-----AHGVTSAPDTRPPAGSTAPPAHGVTS | 160 |
| Qy | 181 | APDRPPALGSTATPPVHNVTYSASGASGASGASTLVHNGTARATYTPPAKSTPASIIPSHSD | 240 |
| Db | 161 | APDRPPALGSTATPPVHNVTYSASGASGASGASTLVHNGTARATYTPPAKSTPESIIPSHSD | 220 |
| Qy | 241 | TPTTLASHSTKTDASTHSTSVPLTSSNHSTSPOLSTGCVFFPLSPHISINIQFNSLED | 300 |
| Db | 221 | TPTTLASHSTKTDASTHSTSVPLTSSNHSTSPOLSTGCVFFPLSPHISINIQFNSLED | 280 |
| Qy | 301 | PSTPYQELORDISEMFLQIQKGGFLGSLNKKPPGVSVMQLAPREGTINVDVETO | 360 |
| Db | 281 | PSTPYQELORDISEMFLQIQKGGFLGSLNKKPPGVSVMQLAPREGTINVDVETO | 340 |
| Qy | 361 | PNQYTRBASRNYLTSDSVSVHPPPFSQSGASGAPMGALTLVTCVLVLAIVLYLA | 420 |
| Db | 341 | PNQYTRBASRNYLTSDSVSVHPPPFSQSGASGAPMGALTLVTCVLVLAIVLYLA | 400 |
| Qy | 421 | LAVOCRRKRYGQLOLFPAPADYHPMSEYVYTHHGRVVRPSSYDSSPPEKYSAGNGSSS | 480 |
| Db | 401 | LAVOCRRKRYGQLOLFPAPADYHPMSEYVYTHHGRVVRPSSYDSSPPEKYSAGNGSSS | 460 |
| Qy | 481 | LSYTPVAATSNLT | 495 |
| Db | 461 | LSYTPVAATSNLT | 475 |

```

RESULT 6
US-60-453-135-8036
? Sequence 8036, Application US/60453135
? GENERAL INFORMATION:
? APPLICANT: CARGILL, Michele
? APPLICANT: IAKOUBOVA, Olga
? TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
? TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
? FILE REFERENCE: CL001456
? CURRENT APPLICATION NUMBER: US/60/453.135
? CURRENT FILING DATE: 2003-03-10
? NUMBER OF SBO ID NOS: 82762
? SOFTWARE: FastSeq for Windows Version 4.0
? SBO ID NO 8036
? LENGTH: 475
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-60-453-135-8036

```

| | | | | |
|-----------------------|--------|--------------------|-------|------------|
| Query Match | 94.2%; | Score 2413; | DB 7; | length 475 |
| Best Local Similarity | 95.6%; | Pred. No. 5.7e-97; | | |

Thu May 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 4

| | Matches | 4/3; Conservative | 0; Mismatches | 2; Indels | 20; Gaps | 1 |
|----|---------|---|---------------|-----------|----------|---|
| Qy | 1 | MPGQGPFLILLTLVTVTSGHASTPCGKERTSATORSSVPSSTKMNVMSTSV | 60 | | | |
| Dh | 1 | MPGQGPFLILLTLVTVTSGHASTPCGKERTSATORSSVPSSTKMNVMSTSV | 60 | | | |
| Qy | 61 | LSHSPGSGSSTTGQDVTLPATPEPSSAATMGQDVSVPVTPALGSGTTPADYTS | 120 | | | |
| Dh | 61 | LSHSPGSGSSTTGQDVTLPATPEPSSAATMGQDVSVPVTPALGSGTTPADYTS | 120 | | | |
| Qy | 121 | ABDNKPAAGSTAPPAHGVTSADTPRPPPGSTAPAAHGVTSADTPRPAAGSTAPPAHGVTS | 180 | | | |
| Dh | 121 | ADNKKPAPGSTAP-----AHGVTSADTPRPAAGSTAPPAHGVTS | 160 | | | |
| Qy | 181 | ABDNKPAAGSTAPPAHGVTSASGSASGSASTLVHNGTSAARATTPAASK3TPPS1PEHNSD | 240 | | | |
| Dh | 161 | ABDNKPAAGSTAPPAHGVTSASGSASGSASTLVHNGTSAARATTPAASK3TPPS1PEHNSD | 220 | | | |
| Qy | 241 | TFTTLASHSTKTDASTHSTVPEPLTSSNHSTSPQLSTGVSPFLSPHISNIDFNSIED | 300 | | | |
| Dh | 221 | TFTTLASHSTKTDASTHSTVPEPLTSSNHSTSPQLSTGVSPFLSPHISNIDFNSIED | 280 | | | |
| Qy | 301 | PESTYQELORISMTQIKYCGFLGSNKRPESSVYVLTAPREGTINVADVTSQ | 360 | | | |
| Dh | 281 | PESTYQELORISMTQIKYCGFLGSNKRPESSVYVLTAPREGTINVADVTSQ | 340 | | | |
| Qy | 361 | PNQYKTEAARNYLTISDVSSVDPPEPSPAGSGAGVGMQALIVAVCYALALVYLLA | 420 | | | |
| Dh | 341 | PNQYKTEAARNYLTISDVSSVDPPEPSPAGSGAGVGMQALIVAVCYALALVYLLA | 400 | | | |
| Qy | 421 | LAVCCCKRNYGGDLPAPRATYHPMSKPTTHHNGYVPESTSPRSYKRYSGAGGSS | 480 | | | |
| Dh | 401 | LAVCCCKRNYGGDLPAPRATYHPMSKPTTHHNGYVPESTSPRSYKRYSGAGGSS | 460 | | | |
| Qy | 481 | LSYTNPAVAATSANL | 495 | | | |
| Dh | 461 | LSYTNPAVAATSANL | 475 | | | |

[illegible]

| | | | | |
|----|-----|-----------------|--|-----|
| Qy | 181 | ADNDPRLASGTA | PPVHNWTSAGSGSAGSASTLVHNWTSABATTTTPASXKSTPSPISPHSHD | 240 |
| Db | 161 | ADNDPRLASGTA | PPVHNWTSAGSGSAGSASTLVHNWTSABATTTTPASKSTPSPISPHSD | 220 |
| Qy | 241 | TEFTTLASHSKTDA | STHSTVPEPLTSSNHSSTPOLSTGVFAFFFLSH1SNLOFNSSLEAD | 300 |
| Db | 221 | TEFTTLASHSKTDA | STHSTVPEPLTSSNHSSTPOLSTGVFAFFFLSH1SNLOFNSSLEAD | 280 |
| Qy | 301 | PSTDYVGEIORDISE | MPLOITKOGGFLSNITFRPGSVVVOULTLAERGTTHNVADVETQ | 360 |
| Db | 281 | PSTDYVGEIORDISE | MPLOITKOGGFLSNITFRPGSVVVOULTLAERGTTHNVADVETQ | 340 |
| Qy | 361 | FNQKCTAASRNLT | LDVSGVSHVPPPSAGSGVPCMGKIALTLVYCYVALMYLTLA | 420 |
| Db | 341 | FNQKCTAASRNLT | LDVSGVSHVPPPSAGSGVPCMGKIALTLVYCYVALMYLTLA | 400 |
| Qy | 421 | LAVOCCKRNOYQGLD | IFPADTYHMSSEYPTVTHORVYVPSSTDSPEKYSAGNCGSS | 480 |
| Db | 401 | LAVOCCKRNOYQGLD | IFPADTYHMSSEYPTVTHORVYVPSSTDSPEKYSAGNCGSS | 460 |
| Qy | 481 | LSYTNPAVAATSANL | 495 | |
| Db | 461 | LSYTNPAVAATSANL | 475 | |

| | | |
|---------------------------|---|--|
| | RESULT 8 | |
| PCT-US02-19669A-311 | Sequence 311, Application PC/TUS0219669A | |
| | GENERAL INFORMATION: | |
| | APPLICANT: Millenium Pharmaceuticals, Inc. et al. | |
| | TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, | |
| | FILE REFERENCE: MRI-0389C | |
| | CURRENT APPLICATION NUMBER: PCT/US02/19669A | |
| | PRIOR FILING DATE: 2001-06-21 | |
| | PRIOR APPLICATION NUMBER: US 60/299,887 | |
| | PRIOR FILING DATE: 2001-06-21 | |
| | PRIOR APPLICATION NUMBER: US 60/301,572 | |
| | PRIOR FILING DATE: 2001-06-27 | |
| | PRIOR APPLICATION NUMBER: US 60/306,501 | |
| | PRIOR FILING DATE: 2001-07-18 | |
| | PRIOR APPLICATION NUMBER: US 60/325,002 | |
| | PRIOR FILING DATE: 2001-09-25 | |
| | PRIOR APPLICATION NUMBER: US 60/362,585 | |
| | PRIOR FILING DATE: 2002-03-05 | |
| | PRIOR APPLICATION NUMBER: US 60/xxx,xxx | |
| | PRIOR FILING DATE: 2002-05-14 | |
| | NUMBER OF SEQ ID NOS: 506 | |
| | SOFTWARE: FastSeq for Windows Version 4.0 | |
| | SEQ ID NO 311 | |
| | LENGTH: 1255 | |
| | TYPE: PRT | |
| | ORGANISM: Homo sapiens | |
| PCT-US02-19669A-311 | | |
| Query Match | 83.1%; Score 2130; DB 1; Length 1255; | |
| Best Local Similarity | 39.0%; Pred. No. 2,1e+84; | |
| Matches 489; Conservative | 1; Mismatches 5; Indels 760; Gaps 1; | |
| QY | 1 MTPGTGPPFLLLLTVLTVVTGSGHAESTPGCEKTSATORSRVSPSTKKNAVSMTSSV 60 | |
| Dc | 1 MTPGTGPPFLLLLTVLTVVTGSGHAESTPGCEKTSATORSRVSPSTKKNAVSMTSSV 60 | |
| QY | 61 LSSHSPPSGSGSTTGQGDVTLAPRTEPAAGSAATWQDVTSVPVTRPALASTTPPAHGVTS 120 | |
| Dc | 61 LSSHSPPSGSGSTTGQGDVTLAPRTEPAAGSAATWQDVTSVPVTRPALASTTPPAHGVTS 120 | |
| QY | 121 APDKKPAGSITAPPAHVGTASPTPRPGSITAFAHGVTSAPDT----- 164 | |
| Dc | 121 APDKKPAGSITAPPAHVGTASPTPRPGSITAFAHGVTSAPDT----- 164 | |
| QY | 165 ----- 164 | |

Db 161 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 960
Qy 201 ASGSASGASTLVHNGTSARATTPPASKSTPPSIPSHSDPTTLASHSTKTDASSTHHS 260
Db 961 ASGSASGASTLVHNGTSARATTPPASKSTPPSIPSHSDPTTLASHSTKTDASSTHHS 1020
Qy 261 TVPPLTSSNNHSTSPOLSTGVSFFLASHISNLOFNSSLEDDSTDTYQELQNDISEMFLQI 320
Db 1021 SVPEPLTSSNNHSTSPOLSTGVSFFLASHISNLOFNSSLEDDSTDTYQELQNDISEMFLQI 1080
Qy 321 YKGGFGLSLNIRPGSVYVQGLTARREGTINHDTGTCNOYKTEASYNLTISDHS 380
Db 1081 YKGGFGLSLNIRPGSVYVQGLTARREGTINHDTGTCNOYKTEASYNLTISDHS 1140
Qy 381 VSHVPPSSASGAGVPGKGTALVYVCVYALAVYLIALAVCCORRNGOGLDIPAR 440
Db 1141 VSHVPPSSASGAGVPGKGTALVYVCVYALAVYLIALAVCCORRNGOGLDIPAR 1200
Qy 441 DTYHMSGYPTTHTGRVYVPPSTDRSPYKVSAGGSSLSSTNPAVAATSAAL 495
Db 1201 DTYHMSGYPTTHTGRVYVPPSTDRSPYKVSAGGSSLSSTNPAVAATSAAL 1255

Result 9

PCT-IL02-00255-2
SEQUENCE 2 Application PC/TL0200255
GENERAL INFORMATION:
APPLICANT: Ramot University Authority for Applied Research
APPLICANT: KRESCHNER, Daniel
APPLICANT: YOBEL-LENER, Merav
APPLICANT: SMORODINSKY, Nechama
TITLE OF INVENTION: Peptides and Antibodies to MUC 1 Proteins
FILE REFERENCE: P-3891-PC
CURRENT APPLICATION NUMBER: PCT/IL02/00255
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: 60/279,408
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-IL02-00255-2
Query Match 83.1%; Score 2130; DB 1; Length 1255;
Best Local Similarity 39.0%; Pred. No. 2.1e-84;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;
Qy 1 MTPGTQSPFFLLLLTLTVTVYVYSGHNASTPGSEKETSATQSSVPSSTERKNAVSMTSY 60
Db 1 MTPGTQSPFFLLLLTLTVTVYVYSGHNASTPGSEKETSATQSSVPSSTERKNAVSMTSY 60
Qy 61 LSSHPGSGSSTTGQDVTTLAPATEPASGSAATMGQDVTSPVTRPALGSTTPPADVTS 120
Db 61 LSSHPGSGSSTTGQDVTTLAPATEPASGSAATMGQDVTSPVTRPALGSTTPPADVTS 120
Qy 121 APDNKAPGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 164
Db 121 APDNKAPGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 161 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS 720

```

Qy 165 ----- 164
Db 721 APDTRPAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APDTRPAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APDTRPAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 APDTRPAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 960
Qy 201 ASGSASGASSTLVNCTSAATTTASKSTPSPISHSNDSPTTASHTTASHTS 260
Db 961 ASGSASGASSTLVNCTSAATTTASKSTPSPISHSNDSPTTASHTTASHTS 1020
Qy 261 TYPPLTSSNHSSTPOLSTGVSPFFLSFHSISNLOFNSSLEDSTDYQOELORDISEMFLQI 320
Db 1021 SVPLTSSNHSSTPOLSTGVSPFFLSFHSISNLOFNSSLEDSTDYQOELORDISEMFLQI 1080
Qy 321 YKGGFLGSLNKKRPGSVVVLTLAFREGTIVHDVETOPNOYKTEASRYNLTISDVS 380
Db 1081 YKGGFLGSLNKKRPGSVVVLTLAFREGTIVHDVETOPNOYKTEASRYNLTISDVS 1140
Qy 381 VSHVPPPSAOSGAGVPGMGIALVLCVVALAIVYLAVALAVOCORRONGOLDIPPAR 440
Db 1141 VSDVPPPSAOSGAGVPGMGIALVLCVVALAIVYLAVALAVOCORRONGOLDIPPAR 1200
Qy 441 DTVHPMSSEPTHTHGRVYPPSTDRSPYKVASGAGSSISYTNPAVAATSANL 495
Db 1201 DTVHPMSSEPTHTHGRVYPPSTDRSPYKVASGAGSSISYTNPAVAATSANL 1255

RESULT 10
US-60-452-680-12838
; Sequence 12838, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGIL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 12838
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-12838

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
Best Local Similarity 80.6%; Pred. No. 6,6e-79;
Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

```

```

Db 161 APDTRPAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 220
Qy 241 TPTTLASHTKTDASSTHSTVPLTSSNHSSTPOLSTGVSPFFLSFHSISNLOFNSSLED 300
Db 221 TPTTLASHTKTDASSTHSTVPLTSSNHSSTPOLSTGVSPFFLSFHSISNLOFNSSLED 280
Qy 301 PSTDYVOELORDISEMFLQIYKGGFLGSLNKKRPGSVVVLTLAFREGTIVHDVETO 360
Db 281 PSTDYVOELORDISEM----- 296
Qy 361 FNOYKTEASRYNLTISDVSHPFPFSAOSGAGVPGMGIALVLCVVALAIVYLA 420
Db 297 ----- 325
Qy 421 LANCORRONGOLDIPPARORTYPMSEPTHTHGRVYPPSTDRSPYKVASGAGSS 480
Db 326 LANCORRONGOLDIPPARORTYPMSEPTHTHGRVYPPSTDRSPYKVASGAGSS 385
Qy 481 LSYTNPAVAATSANL 495
Db 386 LSYTNPAVAATSANL 400

RESULT 11
US-60-453-135-8039
; Sequence 8039, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGIL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 8039
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-8039

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
Best Local Similarity 80.6%; Pred. No. 6,6e-79;
Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

```

Db 297 -----SGAGVGMGIALVLVVCVLAALIVYLLA 325
 Qy 421 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 480
 Db 326 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 385
 Qy 481 LSYNPAVAATSAANL 495
 Db 386 LSYNPAVAATSAANL 400

RESULT 12
 US-60-453-050-8039
 ; Sequence 8039, Application US/60453050
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele
 ; APPLICANT: LUKE, May
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CLO01457
 ; CURRENT APPLICATION NUMBER: US/60/453,050
 ; CURRENT FILING DATE: 2003-03-10
 ; NUMBER OF SEQ ID NOS: 82762
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8039
 ; LENGTH: 400
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-453-050-8039

Query Match 77.8%; Score 1992.5; DB 7; Length 400;
 Beat Local Similarity 80.6%; Pred. No. 6.6e-79;
 Matches 399; Conservative 0; Mismatches 1; Indels 95; Gaps 2;

Qy 1 MTPGTGSPFFLLLLTLVTLVVTGSGHASTPGGKETSAORSVSPSTEKNAVMTSSV 60
 Db 1 MTPGTGSPFFLLLLTLVTLVVTGSGHASTPGGKETSAORSVSPSTEKNAVMTSSV 60
 Qy 61 LSHSPGSGSSTTGGQDVTLAPATBPASGAATWGDVTSVPTPALGSTTPPAHDVTS 120
 Db 61 LSHSPGSGSSTTGGQDVTLAPATBPASGAATWGDVTSVPTPALGSTTPPAHDVTS 120
 Qy 121 APDKRPAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
 Db 121 APDKRPAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
 Qy 181 APDKRPALGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
 Db 181 APDKRPALGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
 Qy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHSINLQPNLSLED 300
 Db 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHSINLQPNLSLED 300
 Qy 301 PSTDYVOELORDISEMFLQIYKGGFLGSLNFKRPSVVVQLTLAPREGTINWHDVETQ 360
 Db 301 PSTDYVOELORDISEMFLQIYKGGFLGSLNFKRPSVVVQLTLAPREGTINWHDVETQ 360
 Qy 361 FNOYKTEASRYNLITSDVSVSHVPPPSAGAGVPGMGIALVLVVCVLAALIVYLLA 420
 Db 361 FNOYKTEASRYNLITSDVSVSHVPPPSAGAGVPGMGIALVLVVCVLAALIVYLLA 420
 Qy 421 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 480
 Db 421 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 480
 Qy 481 LSYNPAVAATSAANL 495
 Db 481 LSYNPAVAATSAANL 400

RESULT 13
 US-60-452-680-12837

; Sequence 12837, Application US/60452680
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele
 ; APPLICANT: GRUPE, Andrew
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CLO01450
 ; CURRENT APPLICATION NUMBER: US/60/452,680
 ; CURRENT FILING DATE: 2003-03-07
 ; NUMBER OF SEQ ID NOS: 116213
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12837
 ; LENGTH: 370
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-452-680-12837

Query Match 71.5%; Score 1831.5; DB 7; Length 370;
 Beat Local Similarity 74.5%; Pred. No. 5.4e-72;
 Matches 369; Conservative 0; Mismatches 1; Indels 125; Gaps 2;

Qy 1 MTPGTGSPFFLLLLTLVTLVVTGSGHASTPGGKETSAORSVSPSTEKNAVMTSSV 60
 Db 1 MTPGTGSPFFLLLLTLVTLVVTGSGHASTPGGKETSAORSVSPSTEKNAVMTSSV 60
 Qy 61 LSHSPGSGSSTTGGQDVTLAPATBPASGAATWGDVTSVPTPALGSTTPPAHDVTS 120
 Db 61 LSHSPGSGSSTTGGQDVTLAPATBPASGAATWGDVTSVPTPALGSTTPPAHDVTS 120
 Qy 121 APDKRPAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
 Db 121 APDKRPAGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 180
 Qy 181 APDKRPALGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
 Db 181 APDKRPALGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPAGSTAPPAHGVTS 240
 Qy 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHSINLQPNLSLED 300
 Db 241 TPTTLASHSTKTDASTHSTHSTVPLTSSNHSTSPQSTGVSPFFLSFHSINLQPNLSLED 300
 Qy 301 PSTDYVOELORDISEMFLQIYKGGFLGSLNFKRPSVVVQLTLAPREGTINWHDVETQ 360
 Db 301 PSTDYVOELORDISEMFLQIYKGGFLGSLNFKRPSVVVQLTLAPREGTINWHDVETQ 360
 Qy 361 FNOYKTEASRYNLITSDVSVSHVPPPSAGAGVPGMGIALVLVVCVLAALIVYLLA 420
 Db 361 FNOYKTEASRYNLITSDVSVSHVPPPSAGAGVPGMGIALVLVVCVLAALIVYLLA 420
 Qy 421 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 480
 Db 421 LAVOCRRKRYGOLDIFPARDTYHPMSEYPTHTHGYYVPPSSTRSPYEVKSAAGGSS 480
 Qy 481 LSYNPAVAATSAANL 495
 Db 481 LSYNPAVAATSAANL 370

RESULT 14
 US-60-453-135-8038
 ; Sequence 8038, Application US/60453135
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele
 ; APPLICANT: IAKUBOVA, Olga
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CLO01456
 ; CURRENT APPLICATION NUMBER: US/60/453,135
 ; CURRENT FILING DATE: 2003-03-10
 ; NUMBER OF SEQ ID NOS: 82762
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8038
 ; LENGTH: 370

Thu May 8 16:14:03 2003

us-09-658-621b-2.rapn

Page 8

```

; TYPE: PRF
; ORGANISM: Homo sapiens
US-60-453-135-8038

```

| | | | | |
|---------------------------|--------|--------------------|-------------|-------------|
| Query Match | 71.5%; | Score 1831.5; | DB 7; | Length 370; |
| Best Local Similarity | 74.5%; | Pred. No. 5.4e-72; | | |
| Matches 369; Conservative | 0; | Mismatches 1; | Indels 125; | Gaps 2; |

[illegible]

```

RESULT 15
US-60-453-050-8038
Sequence 8038, Application US/60453050
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: LUKE, May
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001457
CURRENT APPLICATION NUMBER: US/60/453,050
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8038
LENGTH: 370
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-050-8038

```

| Query Match | Score | 1831.5 | DB 7 | Length | 370 |
|-----------------------|-------|----------------------|-------------------|--------------------|------------|
| Best Local Similarity | 74.5 | Pred. | 0.5 | 4e-72 | |
| Matches | 369 | Conservative | 1 | Indels | 125 |
| | | | | Gaps | 2 |
| Qy | 1 | MPRGSPAPPLLLIYLVYVTS | GHASSTPGCKREKMSA | QSSPSSTKXNAV | WTSYSSV 60 |
| Db | 1 | MPRGSPAPPLLLIYLVYVTS | GHASSTPGCKREKMSA | QSSPSSTKXNAV | WTSYSSV 60 |
| Qy | 6 | LSHSPGSGSITTGQDVTLL | PATPRPSSGAATWGVTS | VPVTPALAGSTTPADVTS | 120 |

| | | | |
|----|-----|--|-----|
| Db | 61 | LSHSPGSGSTTQGDVUTLAPATEPAGSAATWQDVTSPVTRPALSTTTPPAHDVTS | 120 |
| QY | 121 | ABDNKPAAGSTAPPAHGVTASAPDTRPPGSGTAPAAHGVTASAPDRPAPOSTPAPHGVT | 180 |
| Db | 121 | ABDNKPAAGSTAPP-----AHGVTASAPDRPAPOSTPAPHGVT | 160 |
| QY | 181 | ABDNKPALGSTAPPVHNVNTSASGASGASATLVHNGTARATTPPAKSTPTPSISHSD | 240 |
| Db | 161 | ABDNKPALGSTAPPVHNVNTSASGASGASATLVHNGTARATTPPAKSTPTPSISHSD | 220 |
| QY | 241 | TPETTLASHSTKTAGASTHSHSVPLTSSNHSHTSPOLSTGVSFEPFLSFHLSNLOFNSSLED | 300 |
| Db | 221 | TPETTLASHSTKTAGASTHSHSVPLTSSNHSHTSPOLSTGVSFEPFLSFHLSNLOFNSSLED | 280 |
| QY | 301 | PSTDVYOEHLRDISSEMFLOIKYGGGFLGSLNKKFPGSVVOLTLAREGTTNVHDVETQ | 360 |
| Db | 281 | PSTDVYOEHLRDISSE----- | 296 |
| QY | 361 | FQVYKTBASRYNLITSDVSVSHVPFPPSASOGAGVGMGIALVLVCLVALALAVLIA | 420 |
| Db | 297 | ----- | 296 |
| QY | 421 | LAVCCRRKMYGGLDIPARDYTHPMGEYPTHTGRRVPPSTDRSPYEKVASAGNGSS | 480 |
| Db | 297 | -VVCCKRRKMYGGLDIPARDYTHPMGEYPTHTGRRVPPSTDRSPYEKVASAGNGSS | 355 |
| QY | 481 | LSYTNPAVAATSAANL 495 | |
| Db | 356 | LSYTNPAVAATSAANL 370 | |

Search completed: May 1, 2003, 23:21:01
Job time : 110.089 secs

Job time : 110.089 secs

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:10:31 ; Search time 273.036 Seconds
(without alignments)
1168.869 Million cell updates/sec

Title: US-09-658-621b-2

Perfect score: 2562
Sequence: 1 MPTGQSPPLILITLVTV.....NGGSLSTVNPVAVATSNL 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents AA Main:
1: /cgn2_6/prodata/1/paa/US06_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
8: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
9: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
10: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
11: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
12: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
13: /cgn2_6/prodata/1/paa/US08_COMB.pep.*
14: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
15: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
16: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
17: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
18: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
19: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
20: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
21: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
22: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
23: /cgn2_6/prodata/1/paa/US09_COMB.pep.*
24: /cgn2_6/prodata/1/paa/US10_COMB.pep.*
25: /cgn2_6/prodata/1/paa/US10_COMB.pep.*
26: /cgn2_6/prodata/1/paa/US10_COMB.pep.*
27: /cgn2_6/prodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 2562 | 100.0 | 495 | 20 | US-09-658-621-2 |
| 2 | 2516 | 98.2 | 515 | 1 | PCT-US02-07826-212 |
| 3 | 2516 | 98.2 | 515 | 24 | US-10-097-340-212 |
| 4 | 2516 | 98.2 | 515 | 25 | US-10-171-311-156 |
| 5 | 2413 | 94.2 | 475 | 1 | PCT-US01-30151-3 |
| 6 | 2413 | 94.2 | 475 | 20 | US-09-658-621-67 |

| | | | | | | |
|----|--------|------|------|----|---------------------|--------------------|
| 7 | 2282.5 | 89.1 | 508 | 17 | US-09-166-670-20 | Sequence 20, Appl |
| 8 | 2282.5 | 89.1 | 508 | 24 | US-10-057-136-20 | Sequence 20, Appl |
| 9 | 2130 | 83.1 | 1255 | 1 | PCT-US02-19669-311 | Sequence 311, Appl |
| 10 | 2130 | 83.1 | 1255 | 23 | US-09-996-069-10 | Sequence 10, Appl |
| 11 | 2130 | 83.1 | 1255 | 25 | US-10-171-311-158 | Sequence 158, Appl |
| 12 | 2130 | 83.1 | 1255 | 25 | US-10-171-293-311 | Sequence 311, Appl |
| 13 | 1640 | 64.0 | 321 | 1 | PCT-US00-05882-861 | Sequence 861, Appl |
| 14 | 1640 | 64.0 | 321 | 23 | US-09-925-301-861 | Sequence 861, Appl |
| 15 | 1273 | 49.7 | 222 | 26 | US-10-221-279-12252 | Sequence 6112, Ap |
| 16 | 1202.5 | 46.9 | 630 | 26 | US-10-207-655-69 | Sequence 69, Appl |
| 17 | 967.5 | 37.8 | 230 | 27 | US-09-189-987-356 | Sequence 356, Appl |
| 18 | 967.5 | 37.8 | 230 | 27 | US-09-189-987-356 | Sequence 356, Appl |
| 19 | 828 | 32.3 | 168 | 27 | US-09-196-710-6165 | Sequence 6165, Ap |
| 20 | 828 | 32.3 | 168 | 27 | US-09-196-710-6165 | Sequence 6165, Ap |
| 21 | 649 | 25.3 | 129 | 5 | US-08-135-957-3 | Sequence 3, Appl |
| 22 | 649 | 25.3 | 129 | 5 | US-08-135-957-3 | Sequence 3, Appl |
| 23 | 649 | 25.3 | 129 | 6 | US-08-243-685-3 | Sequence 3, Appl |
| 24 | 649 | 25.3 | 129 | 6 | US-08-243-685-3 | Sequence 3, Appl |
| 25 | 465 | 18.1 | 100 | 23 | US-09-965-131-6 | Sequence 3104, Ap |
| 26 | 396 | 15.5 | 72 | 1 | PCT-US01-49923-1 | Sequence 1, Appl |
| 27 | 353 | 13.8 | 71 | 19 | US-09-593-870A-12 | Sequence 12, Appl |
| 28 | 339 | 13.2 | 177 | 20 | US-09-646-028-54 | Sequence 54, Appl |
| 29 | 334 | 13.0 | 172 | 20 | US-09-646-028-54 | Sequence 49, Appl |
| 30 | 265 | 10.3 | 50 | 1 | PCT-US01-49923-6 | Sequence 6, Appl |
| 31 | 252 | 9.8 | 43 | 1 | PCT-US01-49923-5 | Sequence 5, Appl |
| 32 | 251 | 9.8 | 50 | 15 | US-09-194-799C-12 | Sequence 12, Appl |
| 33 | 251 | 9.8 | 51 | 1 | PCT-US97-04493-6 | Sequence 6, Appl |
| 34 | 251 | 9.8 | 51 | 15 | US-09-194-799C-6 | Sequence 6, Appl |
| 35 | 251 | 9.8 | 51 | 15 | US-09-194-799C-6 | Sequence 6, Appl |
| 36 | 248 | 9.7 | 1953 | 27 | US-09-185-361-566 | Sequence 566, App |
| 37 | 248 | 9.7 | 2947 | 27 | US-09-185-361-566 | Sequence 566, App |
| 38 | 248 | 9.7 | 2947 | 27 | US-09-185-361-566 | Sequence 566, App |
| 39 | 243 | 9.5 | 373 | 1 | PCT-US01-14820-617 | Sequence 617, App |
| 40 | 243 | 9.5 | 373 | 1 | PCT-US01-14820-617 | Sequence 617, App |
| 41 | 240 | 9.4 | 377 | 1 | PCT-US00-32990-7 | Sequence 935, App |
| 42 | 240 | 9.4 | 377 | 25 | US-10-149-819-7 | Sequence 7, Appl |
| 43 | 240 | 9.4 | 377 | 25 | US-10-149-819-7 | Sequence 7, Appl |
| 44 | 240 | 9.4 | 377 | 25 | US-10-149-819-7 | Sequence 7, Appl |
| 45 | 234 | 9.1 | 753 | 27 | US-09-185-568-351 | Sequence 351, App |

ALIGNMENTS

RESULT 1
US-09-658-621-2
; Sequence 2, Application US/09658621
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offeinga, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MGC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/09/658, 621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-658-621-2
Query Match 100.0% Score 2562; DB 20; Length 495;

Best Local Similarity 100.0%; Pred. No. 1,6e-160;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPTGQSPFFLLLLTLTVTGGGHAASSTPGGKETSATGSSVPSSTKNAVMTSSV 60
Db 1 MPTGQSPFFLLLLTLTVTGGGHAASSTPGGKETSATGSSVPSSTKNAVMTSSV 60
QY 61 LSSHPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
Db 61 LSSHPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
Db 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
Db 181 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
QY 241 TPTTASHSTKTDASTHSTHSTVPTLSSNHSSTPQSTGVSFFFLSHISNLQFNSSLED 300
Db 241 TPTTASHSTKTDASTHSTHSTVPTLSSNHSSTPQSTGVSFFFLSHISNLQFNSSLED 300
QY 301 PSTDYQELQDISEMFLQIYKQGFGLSNIKERPGSVVQTLAREGTTNVHDVETQ 360
Db 301 PSTDYQELQDISEMFLQIYKQGFGLSNIKERPGSVVQTLAREGTTNVHDVETQ 360
QY 361 PNOYKTEASRYNLTISDVSVSHVPPSAQSGAGVPGMGIALLVCLVALAIVYLIA 420
Db 361 PNOYKTEASRYNLTISDVSVSHVPPSAQSGAGVPGMGIALLVCLVALAIVYLIA 420
QY 421 LAVCCRRKNYGQLDIFPARDTYHPMSEYPTVTHGRVYPPSSSTDSPEYKVSAGNASS 480
Db 421 LAVCCRRKNYGQLDIFPARDTYHPMSEYPTVTHGRVYPPSSSTDSPEYKVSAGNASS 480
QY 481 LSTYNPAAVATSANL 495
Db 481 LSTYNPAAVATSANL 495

RESULT 2
PCT-US02-07826-212
Sequence 212, Application PC/TUS0207826
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceutical, Inc. et al.
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: MRI-030PC
CURRENT APPLICATION NUMBER: PCT/US02/07826
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 212
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-07826-212

Query Match 98.2%; Score 2516; DB 1; Length 515;

Best Local Similarity 95.3%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;

QY 1 MPTGQSPFFLLLLTLTVTGGGHAASSTPGGKETSATGSSVPSSTKNAVMTSSV 60
Db 1 MPTGQSPFFLLLLTLTVTGGGHAASSTPGGKETSATGSSVPSSTKNAVMTSSV 60
QY 61 LSSHPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
Db 61 LSSHPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTRPALGSTTPPAHDVTS 120
QY 121 APDNKPA-----PSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 160
Db 121 APDNKPAAGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 161 APDTRPPAGSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAR 220
Db 161 APDTRPPAGSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAPDNKPAAGSTAPPAHGVTSAR 240
QY 221 ATTPPASKSTPSSISHSSTPPTTLASHSTKTDASTHSTHSTVPTLSSNHSSTPQSTGVS 280
Db 221 ATTPPASKSTPSSISHSSTPPTTLASHSTKTDASTHSTHSTVPTLSSNHSSTPQSTGVS 300
QY 281 SFFFLSHISNLQFNSSLEDSTDYQELQDISEMFLQIYKQGFGLSNIKERPGSVV 340
Db 281 SFFFLSHISNLQFNSSLEDSTDYQELQDISEMFLQIYKQGFGLSNIKERPGSVV 360
QY 341 VOLTLAREGTTNVHDVETQFNOYKTEASRYNLTISDVSVSHVPPSAQSGAGVPGMG 400
Db 341 VOLTLAREGTTNVHDVETQFNOYKTEASRYNLTISDVSVSHVPPSAQSGAGVPGMG 420
QY 401 IALLVCLVALAIVYLIALAVCCRRKNYGQLDIFPARDTYHPMSEYPTVTHGRVY 460
Db 401 IALLVCLVALAIVYLIALAVCCRRKNYGQLDIFPARDTYHPMSEYPTVTHGRVY 480
QY 461 PSTDSPEYKVSAGNASSISTYNPAAVATSANL 495
Db 461 PSTDSPEYKVSAGNASSISTYNPAAVATSANL 515

RESULT 3
US-10-097-340-212
Sequence 212, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAWATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel B. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26

Thu May 3 16:14:02 2003

us-09-658-621b-2.rapm

Page 3

```
/ PRIOR APPLICATION NUMBER: 60/311,732
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: 60/325,102
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: 60/323,580
/ PRIOR FILING DATE: 2001-09-19
/ NUMBER OF SEQ ID NOS: 363
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 212
/ LENGTH: 515
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-097-340-212
```

```
Query Match 98.2%; Score 2516; DB 24; Length 515;
Best Local Similarity 95.5%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;
```

```
OY 1 MTGCTGSPFFLLLLTLVTVTGSGHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
DB 1 MTGCTGSPFFLLLLTLVTVTGSGHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
OY 61 LSSHSBGSSSTTGQODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPPAHVTS 120
DB 61 LSSHSBGSSSTTGQODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPPAHVTS 120
OY 121 APDNKA-----PGSTAPPAGVTSAPDTRPPPGSTAPAAHGVTS 160
DB 121 APDNKA-----PGSTAPPAGVTSAPDTRPPPGSTAPAAHGVTS 160
OY 161 APDTRAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTS 180
DB 161 APDTRAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTS 180
OY 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 280
DB 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 280
OY 241 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 300
DB 241 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 300
OY 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 340
DB 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 340
OY 301 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 360
DB 301 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 360
OY 341 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 400
DB 341 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 400
OY 361 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 420
DB 361 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 420
OY 401 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 460
DB 401 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 460
OY 421 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 480
DB 421 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 480
OY 461 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
DB 461 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
OY 481 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 515
DB 481 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 515
```

```
RESULT 4
US-10-171-311-156
/ Sequence 156, Application US/10171311
/ GENERAL INFORMATION:
/ APPLICANT: Schlegel, Robert
/ APPLICANT: Chen, Yan
/ APPLICANT: Zhao, Xumei
/ APPLICANT: Monahan, John
/ APPLICANT: Glatz, Karen
/ APPLICANT: Gannavarapu, Manjula
/ APPLICANT: Hoersb, Sebastian
/ TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
/ TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
/ FILE REFERENCE: MRI-035
/ CURRENT APPLICATION NUMBER: US/10/171,311
/ CURRENT FILING DATE: 2002-06-12
```

```
/ PRIOR APPLICATION NUMBER: US 60/298,159
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/298,155
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/335,936
/ PRIOR FILING DATE: 2001-11-14
/ NUMBER OF SEQ ID NOS: 238
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 156
/ LENGTH: 515
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-171-311-156
```

```
Query Match 98.2%; Score 2516; DB 25; Length 515;
Best Local Similarity 95.5%; Pred. No. 1,9e-157;
Matches 492; Conservative 0; Mismatches 3; Indels 20; Gaps 1;
```

```
OY 1 MTGCTGSPFFLLLLTLVTVTGSGHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
DB 1 MTGCTGSPFFLLLLTLVTVTGSGHASTPGGKETSATQRSVSPSTKNAVSMTSV 60
OY 61 LSSHSBGSSSTTGQODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPPAHVTS 120
DB 61 LSSHSBGSSSTTGQODVTLAPATEPAGSAAATGQDVTSPVTRPALGSTPPAHVTS 120
OY 121 APDNKA-----PGSTAPPAGVTSAPDTRPPPGSTAPAAHGVTS 160
DB 121 APDNKA-----PGSTAPPAGVTSAPDTRPPPGSTAPAAHGVTS 160
OY 161 APDTRAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTS 180
DB 161 APDTRAPGSTAPPAHGVTSAPDTRPPPGSTAPPAHGVTSAPDTRPPPGSTAPAAHGVTS 180
OY 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 280
DB 221 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 280
OY 241 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 300
DB 241 ATTPPASKTSPSPISPSHSDPTTLASSTKTDASTHSTVPPPLTSSNMTSPOLSTGV 300
OY 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 340
DB 281 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 340
OY 301 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 360
DB 301 SFFFLSFHNSLNQFNSSLEDPSTDYQELQDISEMFLQIYKGGFLGLSNIKFRGSAV 360
OY 341 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 400
DB 341 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 400
OY 361 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 420
DB 361 VOLTLAFREGTINHADVETOQNOYKTEAASRYNLTISDVSVDVDFPPPSAQSAGVPGMG 420
OY 401 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 460
DB 401 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 460
OY 421 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 480
DB 421 IALLVVCVLAIVLAIVLIALAVCCCRKNYGOLDIFPARDTYHMSRYPTHTHGRVVP 480
OY 461 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
DB 461 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
OY 481 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 515
DB 481 PSSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 515
```

```
RESULT 5
PCT-US01-30151-3
/ Sequence 3, Application PC/TUS0130151
/ GENERAL INFORMATION:
/ APPLICANT: Genasance Pharmaceutical, Inc.
/ APPLICANT: Koshiy, Beena
/ APPLICANT: Koshiy, Beena
/ TITLE OF INVENTION: HAPLOTYPES OF THE MUC1 GENE
/ FILE REFERENCE: MUC1 MMH-1408PCT
/ CURRENT APPLICATION NUMBER: PCT/US01/30151
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: 60/236,113
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 475
/ TYPE: PRT
```

Thu May 8 16:14:02 2003

us-09-658-621b-2.rapm

Page 4

```

; ORGANISM: Homo sapiens
PCT-US01-30151-3

Query Match          94.2%; Score 2413; DB 1; Length 475;
Best Local Similarity 95.6%; Pred. No. 1,1e-150;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60
DB 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60
QY 61 LSSSPSGSSTTQGGDVTLPATERPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
DB 61 LSSSPSGSSTTQGGDVTLPATERPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
DB 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 181 APDNRPALGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
DB 181 APDNRPALGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHSLSLQFNSSLED 300
DB 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHSLSLQFNSSLED 300
QY 301 PSTDYOELOORDISEMFLQIKQGGFLGSLNKRPPGSVVVQLTLAFREGTINVDVETQ 360
DB 301 PSTDYOELOORDISEMFLQIKQGGFLGSLNKRPPGSVVVQLTLAFREGTINVDVETQ 360
QY 361 FNQKTEAARVNLITSDVSYSVHPPPPSAQSAGVPGWGIALLVYCVLAVALIYVLLA 420
DB 361 FNQKTEAARVNLITSDVSYSVHPPPPSAQSAGVPGWGIALLVYCVLAVALIYVLLA 420
QY 421 LAVOCGRKNYQGLDIPPARDTYHMSSEYPTHTGRVYPPSSSTDSPYEKVSANGCGSS 480
DB 421 LAVOCGRKNYQGLDIPPARDTYHMSSEYPTHTGRVYPPSSSTDSPYEKVSANGCGSS 480
QY 481 LSYNPAVAATSANL 495
DB 481 LSYNPAVAATSANL 495
DB 461 LSYNPAVAATSANL 475

RESULT 6
US-09-658-621-67
; Sequence 67, Application US/09658621
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Oettinger, Rieunk
; APPLICANT: Weisler, Cornelia Johanna Maria
; APPLICANT: Mettes, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: WOC-I derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/09/658, 621
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 67
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-658-621-67

Query Match          94.2%; Score 2413; DB 20; Length 475;

```

```

Best Local Similarity 95.6%; Pred. No. 1,1e-150;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;

QY 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60
DB 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60
QY 61 LSSSPSGSSTTQGGDVTLPATERPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
DB 61 LSSSPSGSSTTQGGDVTLPATERPASGAATWGDVTSVPTRPALGSTTPPAHGVTS 120
QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
DB 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 180
QY 181 APDNRPALGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
DB 181 APDNRPALGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTSAPDTRPPGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHSLSLQFNSSLED 300
DB 241 TPTTLASHTKTDASTHSTVPLTSSNHSTSPOLSTGVSEFFLSFHSLSLQFNSSLED 300
QY 301 PSTDYOELOORDISEMFLQIKQGGFLGSLNKRPPGSVVVQLTLAFREGTINVDVETQ 360
DB 301 PSTDYOELOORDISEMFLQIKQGGFLGSLNKRPPGSVVVQLTLAFREGTINVDVETQ 360
QY 361 FNQKTEAARVNLITSDVSYSVHPPPPSAQSAGVPGWGIALLVYCVLAVALIYVLLA 420
DB 361 FNQKTEAARVNLITSDVSYSVHPPPPSAQSAGVPGWGIALLVYCVLAVALIYVLLA 420
QY 421 LAVOCGRKNYQGLDIPPARDTYHMSSEYPTHTGRVYPPSSSTDSPYEKVSANGCGSS 480
DB 421 LAVOCGRKNYQGLDIPPARDTYHMSSEYPTHTGRVYPPSSSTDSPYEKVSANGCGSS 480
QY 481 LSYNPAVAATSANL 495
DB 481 LSYNPAVAATSANL 495
DB 461 LSYNPAVAATSANL 475

RESULT 7
US-09-366-670-20
; Sequence 20, Application US/09366670
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUPE, DONALD
; APPLICANT: PANICALI, DENNIS
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST WOC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/09/366, 670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; EARLIER FILING DATE: 1998-02-24
; EARLIER APPLICATION NUMBER: 60/038, 253
; EARLIER FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-366-670-20

Query Match          89.1%; Score 2282.5; DB 17; Length 508;
Best Local Similarity 89.2%; Pred. No. 4.9e-142;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps 3;

QY 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60
DB 1 MTPGOSPPFLILLITLVLTVTGSGHASTPGGKETSATORSVSSSTEKNAISMSSV 60

```

```
OY 52 NAVSMTSSVLSHSPGSGSSTTQGDVTLAPATEPAGSAAATWGDVTSVPTPALGST 111
    |||||
DB 61 NAVSMTS--L1SNSGVRGSTAPPAHGVTSAPDTRPAGSTPAHAGVTSAPTRPAGST 118
    |||||
OY 112 TTPADVTSAPDNKPAAGSTAPPAHGVTSAPDTR-----PPGSTAPPAHGVTSAPDTR 165
    |||||
DB 119 APHAGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 178
    |||||
OY 166 PARGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 225
    |||||
DB 179 PARGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 238
    |||||
OY 226 ASKSTPPGTPSHSSTPTTTLASHSTKTDASSTHSTVPTLSSNHSSTPOLSTGVSTFFEL 285
    |||||
DB 239 ASKSTPPGTPSHSSTPTTTLASHSTKTDASSTHSTVPTLSSNHSSTPOLSTGVSTFFEL 298
    |||||
OY 286 SFHISNLOPNSLEPSTDYQELORDISEMFLQIYKOGGFLGSLNIFKFGSVVQULTL 345
    |||||
DB 299 SFHISNLOPNSLEPSTDYQELORDISEMFLQIYKOGGFLGSLNIFKFGSVVQULTL 358
    |||||
OY 346 APREGTINHVDETQFNOYKTEPAASRYNLISDVSUSVPPPSAGSAGVPGMGJALLL 405
    |||||
DB 359 APREGTINHVDETQFNOYKTEPAASRYNLISDVSUSVPPPSAGSAGVPGMGJALLL 418
    |||||
OY 406 LVCVVALAIVYLALAVCCCRKRYGOLDIFPARDTYHPMSEYPTHTHGRVPPSSTD 465
    |||||
DB 419 LVCVVALAIVYLALAVCCCRKRYGOLDIFPARDTYHPMSEYPTHTHGRVPPSSTD 478
    |||||
OY 466 RSPYKVSAGNGSSLSSTNPAVAATSNL 495
    |||||
DB 479 RSPYKVSAGNGSSLSSTNPAVAATSNL 508
    |||||
```

```
RESULT 8
US-10-057-136-20
; Sequence 20, Application US/10057136
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUPE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20
```

```
Query Match 89.1%; Score 2292.5; DB 24; Length 508;
Best Local Similarity 89.2%; Pred. No. 4,9e-142;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps 3;
```

```
OY 112 TTPADVTSAPDNKPAAGSTAPPAHGVTSAPDTR-----PPGSTAPPAHGVTSAPDTR 165
    |||||
DB 119 APHAGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 178
    |||||
OY 166 PARGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 225
    |||||
DB 179 PARGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 238
    |||||
OY 226 ASKSTPPGTPSHSSTPTTTLASHSTKTDASSTHSTVPTLSSNHSSTPOLSTGVSTFFEL 285
    |||||
DB 239 ASKSTPPGTPSHSSTPTTTLASHSTKTDASSTHSTVPTLSSNHSSTPOLSTGVSTFFEL 298
    |||||
OY 286 SFHISNLOPNSLEPSTDYQELORDISEMFLQIYKOGGFLGSLNIFKFGSVVQULTL 345
    |||||
DB 299 SFHISNLOPNSLEPSTDYQELORDISEMFLQIYKOGGFLGSLNIFKFGSVVQULTL 358
    |||||
OY 346 APREGTINHVDETQFNOYKTEPAASRYNLISDVSUSVPPPSAGSAGVPGMGJALLL 405
    |||||
DB 359 APREGTINHVDETQFNOYKTEPAASRYNLISDVSUSVPPPSAGSAGVPGMGJALLL 418
    |||||
OY 406 LVCVVALAIVYLALAVCCCRKRYGOLDIFPARDTYHPMSEYPTHTHGRVPPSSTD 465
    |||||
DB 419 LVCVVALAIVYLALAVCCCRKRYGOLDIFPARDTYHPMSEYPTHTHGRVPPSSTD 478
    |||||
OY 466 RSPYKVSAGNGSSLSSTNPAVAATSNL 495
    |||||
DB 479 RSPYKVSAGNGSSLSSTNPAVAATSNL 508
    |||||
```

```
RESULT 9
PCT-US02-19669-311
; Sequence 311, Application PC/TUS0219669
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc. et al.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038PC
; CURRENT APPLICATION NUMBER: PCT/US02/19669
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-19669-311
```

```
Query Match 83.1%; Score 2130; DB 1; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1.8e-131;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;
```

Db 121 APDKKPAAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 900
Qy 165 ----- 164
Db 901 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 960
Qy 201 ASGASGASGATTVHNGTSARATTPASKSTPPSPSHSSTPTTLASHTKSTASSTHS 260
Db 961 ASGASGASGATTVHNGTSARATTPASKSTPPSPSHSSTPTTLASHTKSTASSTHS 1020
Qy 261 TVPPLTSSMNSTPOLSTGVFFFLSFHISNLQFNSSLEDPSTDYOLQDISEMFLQI 320
Db 1021 SVPLTSSMNSTPOLSTGVFFFLSFHISNLQFNSSLEDPSTDYOLQDISEMFLQI 1080
Qy 321 YKGGGFTGLSNIKFRPGSVVVLTLAFREGTINVDVETQFNQYKTEASRYNLTSQVS 380
Db 1081 YKGGGFTGLSNIKFRPGSVVVLTLAFREGTINVDVETQFNQYKTEASRYNLTSQVS 1140
Qy 381 VSHVPPFPFASGAGVPGMGIALVLVLCVLVALIIVLILALAVCCCRKXNGOLDIFPAR 440
Db 1141 VSDVPPFPFASGAGVPGMGIALVLVLCVLVALIIVLILALAVCCCRKXNGOLDIFPAR 1200
Qy 441 DTYHPMSSEPTVHTGRTVPPSSTDRSPYEKVSAGNGSSLSYTNPAVAANSNL 495

Db 1201 DTYHPMSSEPTVHTGRTVPPSSTDRSPYEKVSAGNGSSLSYTNPAVAANSNL 1255
RESULT 10
US-09-996-069-10
; Sequence 10, Application US/09996069
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996, 069
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10
Query Match 83.1%; Score 2130; DB 23; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1.8e-131;
Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;
Qy 1 MTPGTQSPFFLLLLTLVLTVTYTGSGHASTPGGEKETSATQSSVPSSTERNKAVSMTSV 60
Db 1 MTPGTQSPFFLLLLTLVLTVTYTGSGHASTPGGEKETSATQSSVPSSTERNKAVSMTSV 60
Qy 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPAHDVTS 120
Db 61 LSHSPGSGSSTTQGDVTLAPATEPASGSAATWGDVTSVPTTRPALGSTTPPAHDVTS 120
Qy 121 APDNRPAAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 164
Db 121 APDNRPAAGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 164
Qy 165 ----- 164
Db 165 ----- 164
Qy 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
Db 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 720

Qy 165 ----- 164
Db 721 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 960
Qy 201 AAGSAGSASTLVHNGTSARATTPPASKSTPPSI9SHSDPTTLASHSTKTDASSTHS 260
Db 961 AAGSAGSASTLVHNGTSARATTPPASKSTPPSI9SHSDPTTLASHSTKTDASSTHS 1020
Qy 261 TVPPLTSNHSSTPOLSTGVSPFLSFHISNLOFNSSLSDPTDYOELOORDISEMFLQI 320
Db 1021 SVPLTSSNHSSTPOLSTGVSPFLSFHISNLOFNSSLSDPTDYOELOORDISEMFLQI 1080
Qy 321 YKGGFELGSIKPRGSGVYVOLTLAEBGTINMVDVETQNOYKTBASRYNLTIDVS 360
Db 1081 YKGGFELGSIKPRGSGVYVOLTLAEBGTINMVDVETQNOYKTBASRYNLTIDVS 1140
Qy 381 VSHVPPPSAAGSAGVPGWGLALVLCVVALATYVILALAVQCCRRKNGOLDIFPAR 440
Db 1141 VSHVPPPSAAGSAGVPGWGLALVLCVVALATYVILALAVQCCRRKNGOLDIFPAR 1200
Qy 441 DTYHMSGYPTHTHGRVPPSTDRSPRYEKVAGGSSLTSTNPAVAATNANI 495
Db 1201 DTYHMSGYPTHTHGRVPPSTDRSPRYEKVAGGSSLTSTNPAVAATNANI 1255

RESULT 11
US-10-171-311-158
; Sequence 158, Application US/10171311
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakazi, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerster, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: OF CERIVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Query Match 83.1%; Score 2130; DB 25; Length 1255;
Best Local Similarity 39.0%; Pred. No. 1.8e-13;
Matches 489, Conservative 1; Mismatches 5; Indels 760; Gaps 1;

Qy 1 MTPGTSQFFFLILITLTVTGTSGSAGASTRGKETSATORSVPSSTERNAVSMSTSV 60
Db 1 MTPGTSQFFFLILITLTVTGTSGSAGASTRGKETSATORSVPSSTERNAVSMSTSV 60
Qy 61 LSHSFGSGSSTTQGVDTTLAPATERPAGSAAITWGDVTSVPTPAPAGSTPAPADYTS 120
Db 61 LSHSFGSGSSTTQGVDTTLAPATERPAGSAAITWGDVTSVPTPAPAGSTPAPADYTS 120
Qy 121 APDNKPRAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 164
Db 121 APDNKPRAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 180
Qy 165 ----- 164
Db 181 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 240
Qy 165 ----- 164
Db 241 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 300
Qy 165 ----- 164
Db 301 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 360
Qy 165 ----- 164
Db 361 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 420
Qy 165 ----- 164
Db 421 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 480
Qy 165 ----- 164
Db 481 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 540
Qy 165 ----- 164
Db 541 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 600
Qy 165 ----- 164
Db 601 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 660
Qy 165 ----- 164
Db 661 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 AADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTSADTRPAGSTAPPAHGVTS 960
Qy 201 AAGSAGSASTLVHNGTSARATTPPASKSTPPSI9SHSDPTTLASHSTKTDASSTHS 260
Db 961 AAGSAGSASTLVHNGTSARATTPPASKSTPPSI9SHSDPTTLASHSTKTDASSTHS 1020
Qy 261 TVPPLTSNHSSTPOLSTGVSPFLSFHISNLOFNSSLSDPTDYOELOORDISEMFLQI 320
Db 1021 SVPLTSSNHSSTPOLSTGVSPFLSFHISNLOFNSSLSDPTDYOELOORDISEMFLQI 1080

| | | | |
|----|------|--|------|
| Qy | 321 | YAGGGFGLSNIFRGSGVVOVLTLLAREGTLNVHDTVQCNQKTEASRNLITSDVS | 380 |
| Db | 1081 | YAGGGFGLSNIFRGSGVVOVLTLLAREGTLNVHDTVQCNQKTEASRNLITSDVS | 1140 |
| Qy | 381 | VSRHVPFSAAGSAGVGPGMGIALLVCLVALATVYLTALVCCRRNRYGOLDIFEPAR | 440 |
| Db | 1141 | VSRHVPFSAAGSAGVGPGMGIALLVCLVALATVYLTALVCCRRNRYGOLDIFEPAR | 1200 |
| Qy | 441 | PTVHMSSTPTVTHGVYVPESTDRPEPKYSAGNGGSSLTSTYNPAVAASATL | 495 |
| Db | 1201 | PTVHMSSTPTVTHGVYVPESTDRPEPKYSAGNGGSSLTSTYNPAVAASATL | 1255 |

RESULT 12
US-10-177-293-311

```

Sequence 311, Application US/10177293
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Giact, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamathkar, Shubhang
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Puzrcai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/799,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 311
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-311

```

| Query Match | 83.1% | Score 2130 | DB 25 | Length 1255 |
|-----------------------|----------------|--|------------|-------------|
| Best Local Similarity | 39.0% | Pred No. 1.9e-131 | | |
| Matches 489 | Conservative 1 | Mismatches 3 | Indels 760 | Gaps 1 |
| QY | 1 | MPFGSGPFLLLLTUTVTVTGSGHAASSTPGGKEKTSATORSVPSSTERNASVMTSSV | 60 | |
| Db | 1 | MPFGSGPFLLLLTUTVTVTGSGHAASSTPGGKEKTSATORSVPSSTERNASVMTSSV | 60 | |
| QY | 61 | LSHSGPSGSGSTTGCGDVTLLPATPEPAGSGSAATGCDVTSVPVTPRPAHGVTSPAHADVTS | 120 | |
| Db | 61 | LSHSGPSGSGSTTGCGDVTLLPATPEPAGSGSAATGCDVTSVPVTPRPAHGVTSPAHADVTS | 120 | |
| QY | 121 | APDNKPAAGSTAPPAHAGVTSAPDRPPPGSGTAPPAHGVTSAPDT----- | 164 | |

| | | | |
|----|------|---|------|
| Dd | 121 | ADNDTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 180 |
| Qy | 165 | ----- | 164 |
| Dd | 181 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 240 |
| Qy | 165 | ----- | 164 |
| Dd | 241 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 300 |
| Qy | 165 | ----- | 164 |
| Dd | 301 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 360 |
| Qy | 165 | ----- | 164 |
| Dd | 361 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 420 |
| Qy | 165 | ----- | 164 |
| Dd | 421 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 480 |
| Qy | 165 | ----- | 164 |
| Dd | 481 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 540 |
| Qy | 165 | ----- | 164 |
| Dd | 541 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 600 |
| Qy | 165 | ----- | 164 |
| Dd | 601 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 660 |
| Qy | 165 | ----- | 164 |
| Dd | 661 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 720 |
| Qy | 165 | ----- | 164 |
| Dd | 721 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 780 |
| Qy | 165 | ----- | 164 |
| Dd | 781 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 840 |
| Qy | 165 | ----- | 164 |
| Dd | 841 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 900 |
| Qy | 165 | ----- | 164 |
| Dd | 901 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 960 |
| Qy | 201 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 260 |
| Dd | 961 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 1020 |
| Qy | 261 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 320 |
| Dd | 1021 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 1080 |
| Qy | 321 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 380 |
| Dd | 1081 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 1140 |
| Qy | 381 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 440 |
| Dd | 1141 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 1200 |
| Qy | 441 | ADDTTPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTS | 495 |

```
Db 1201 DTYHMSERYPTVTHGRVYPPSSTDRSPYKESAGNGSSLSSTYNPAVAASANL 1255
RESULT 13
PCT-US00-05882-861
; Sequence 861, Application PC/TUS0005882
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Human Cancer Associated Gene Sequences and Polypeptides
; FILE REFERENCE: PA106PCT
; CURRENT APPLICATION NUMBER: PCT/US00/05882
; EARLIER FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 861
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05882-861

Query Match          64.0%; Score 1640; DB 1; Length 321;
Best Local Similarity 99.1%; Pred. No. 7,6e-100;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query 175 AHGVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 234
Db 1 AHGVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 60
Qy 235 PSHSDPTPTTASHSTKTDASTHSTVPLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 294
Db 61 PSHSDPTPTTASHSTKTDASTHSTVPLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 120
Qy 295 NSSLBDSPTDYQELQDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLAFREGTINV 354
Db 121 NSSLBDSPTDYQELQDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLAFREGTINV 180
Qy 355 HDVETQFNQYKTEASRYNLTISDVSVSHVPPFSAQSGAGVPGMGIALLVLCYVALA 414
Db 181 HDVETQFNQYKTEASRYNLTISDVSVSDVPFSAQSGAGVPGMGIALLVLCYVALA 240
Qy 415 IYVLLALAVCCCRKKNYGOLDIFPARDTYHPMSEYPTVTHGRVYPPSSTDRSPYKESA 474
Db 241 IYVLLALAVCCCRKKNYGOLDIFPARDTYHPMSEYPTVTHGRVYPPSSTDRSPYKESA 300
Qy 475 GNGSSLSSTYNPAVAATSANL 495
Db 301 GNGSSLSSTYNPAVAATSANL 321

RESULT 14
US-09-925-301-861
; Sequence 861, Application US/09925301
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 861
; LENGTH: 321
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match          64.0%; Score 1640; DB 23; Length 321;
Best Local Similarity 99.1%; Pred. No. 7,6e-100;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Query 175 AHGVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 234
Db 1 AHGVTSAPDNRPALGSTAPPVNNTVSAGSAGSASTLVHNGTSARATTTTPASKSTPSSI 60
Qy 235 PSHSDPTPTTASHSTKTDASTHSTVPLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 294
Db 61 PSHSDPTPTTASHSTKTDASTHSTVPLTSSNHSSTSPOLSTGVSEFFLSTHLSNQF 120
Qy 295 NSSLBDSPTDYQELQDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLAFREGTINV 354
Db 121 NSSLBDSPTDYQELQDISEMFLQIYKGGFLGLSNIKFRPGSVVQTLAFREGTINV 180
Qy 355 HDVETQFNQYKTEASRYNLTISDVSVSHVPPFSAQSGAGVPGMGIALLVLCYVALA 414
Db 181 HDVETQFNQYKTEASRYNLTISDVSVSDVPFSAQSGAGVPGMGIALLVLCYVALA 240
Qy 415 IYVLLALAVCCCRKKNYGOLDIFPARDTYHPMSEYPTVTHGRVYPPSSTDRSPYKESA 474
Db 241 IYVLLALAVCCCRKKNYGOLDIFPARDTYHPMSEYPTVTHGRVYPPSSTDRSPYKESA 300
Qy 475 GNGSSLSSTYNPAVAATSANL 495
Db 301 GNGSSLSSTYNPAVAATSANL 321

RESULT 15
US-10-221-279-12252
; Sequence 12252, Application US/10221279
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-046
; CURRENT APPLICATION NUMBER: US/10/221,279
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; NUMBER OF SEQ ID NOS: 12360
; SOFTWARE: Cetus
; SEQ ID NO 12252
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-221-279-12252

Query Match          49.7%; Score 1273; DB 26; Length 292;
Best Local Similarity 54.9%; Pred. No. 1.2e-75;
Matches 272; Conservative 0; Mismatches 1; Indels 22; Gaps 1;

Query 1 MTGTGSPPLLLILTLVTVGSGHASTTGAKETSAVQSSVPSSTERKANVSWTSSV 60
Db 20 MTGTGSPPLLLILTLVTVGSGHASTTGAKETSAVQSSVPSSTERKANVSWTSSV 72
Qy 61 LSHSPGSGSSSTTGQDVTLPATPEPASGAATWGDVTSVPTVRPALSGSTTPPAHDVTS 120
Db 73 ----- 72
Qy 121 APDNKPAGSTAPPAHGVTSADPTRPPGSTAPPAHGVTSADPTRPPAGSTAPPAHGVTS 180
Db 73 ----- 72
```

Thu May 8 16:14:02 2003

us-09-658-621b-2.ram

Page 10

```
Qy 181 APDNRPALGSTAPVHVNTVSAGSAGSASTLVHNGTSARATTPPASKSTPPSIPSHSD 240
Db 73 ----- 72
Qy 241 TPTTLASHSTKTDASSHTSHSTVPLTSSNHSTSPOLSTGVSPFPLSFHISNLQFNSLED 300
Db 73 -----LSTGVSPFPLSFHISNLQFNSLED 97
Qy 301 PSTDYVOELQORDISEMFLQIYKOGFLGLSNIKFRPGSVVWOLTLAFREGTINVHDYETQ 360
Db 98 PSTDYVOELQORDISEMFLQIYKOGFLGLSNIKFRPGSVVWOLTLAFREGTINVHDYETQ 157
Qy 361 FNOYKTEAASRYNLTI SDVSVSHVPPFSAQSGAGVPGWGIALLVYCVLVALAIVYLIA 420
Db 158 FNOYKTEAASRYNLTI SDVSVSDVPPFSAQSGAGVPGWGIALLVYCVLVALAIVYLIA 217
Qy 421 LAVQCCRRKRYGOLDIPPARPTYHPMSEYPTHTHGRYPSPSTDSPPEKVSAGNGSS 480
Db 218 LAVQCCRRKRYGOLDIPPARPTYHPMSEYPTHTHGRYPSPSTDSPPEKVSAGNGSS 277
Qy 481 LSYTNPAVAATSANL 495
Db 278 LSYTNPAVAATSANL 292
```

Search completed: May 1, 2003, 23:19:09
Job time : 281.036 secs

Thu May 8 16:14:01 2003

us-09-658-621b-2.rapb

Page 1

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:12:56 ; Search time 31.4286 Seconds
(without alignments)
1359.015 Million cell updates/sec

Title: US-09-658-621B-2
Perfect score: 2562
Sequence: 1 MPTQSGPFFLLLT/LTVLTV.....NGSSSLSTYNPAAVATSNL 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 8628695 residues
Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|--------------------|--------------------|
| 1 | 2282.5 | 89.1 | 508 | 9 | US-10-057-136-20 | Sequence 20, App1 |
| 2 | 2130 | 83.1 | 1255 | 9 | US-09-996-069-10 | Sequence 10, App1 |
| 3 | 1640 | 64.0 | 321 | 10 | US-09-925-301-861 | Sequence 6, App1 |
| 4 | 465 | 18.1 | 100 | 9 | US-09-965-131-6 | Sequence 7, App1 |
| 5 | 240 | 9.4 | 377 | 9 | US-10-149-819-7 | Sequence 34, App1 |
| 6 | 234 | 9.1 | 1162 | 10 | US-09-745-008-34 | Sequence 108, App1 |
| 7 | 233.5 | 9.1 | 1367 | 10 | US-09-801-368-108 | Sequence 105, App1 |
| 8 | 230 | 9.0 | 554 | 9 | US-09-996-069-3 | Sequence 106, App1 |
| 9 | 230 | 9.0 | 957 | 10 | US-10-025-217-1065 | Sequence 106, App1 |
| 10 | 230 | 9.0 | 957 | 10 | US-09-922-281-1065 | Sequence 106, App1 |
| 11 | 216.5 | 8.5 | 5179 | 10 | US-10-025-380-1068 | Sequence 106, App1 |
| 12 | 216.5 | 8.5 | 5179 | 10 | US-09-922-217-1068 | Sequence 106, App1 |
| 13 | 216.5 | 8.5 | 5179 | 10 | US-09-833-263-1068 | Sequence 106, App1 |
| 14 | 216.5 | 8.5 | 5179 | 10 | US-09-996-069-7 | Sequence 2, App1 |
| 15 | 215 | 8.4 | 51 | 9 | US-09-996-069-2 | Sequence 2, App1 |
| 16 | 215 | 8.4 | 51 | 9 | US-09-996-069-2 | Sequence 2, App1 |
| 17 | 206.5 | 8.1 | 605 | 10 | US-09-801-368-108 | Sequence 428, App1 |
| 18 | 205 | 8.0 | 40 | 9 | US-09-996-069-9 | Sequence 9, App1 |
| 19 | 205 | 8.0 | 46 | 9 | US-09-996-069-9 | Sequence 9, App1 |

| | | | | | | |
|----|-------|-----|------|----|---------------------|--------------------|
| 20 | 205 | 8.0 | 1060 | 10 | US-09-955-909-2 | Sequence 2, App1 |
| 21 | 192 | 7.5 | 941 | 12 | US-10-124-557-14 | Sequence 14, App1 |
| 22 | 192 | 7.5 | 1022 | 12 | US-10-124-557-84 | Sequence 84, App1 |
| 23 | 192 | 7.5 | 1038 | 12 | US-10-124-557-74 | Sequence 74, App1 |
| 24 | 192 | 7.5 | 1049 | 12 | US-10-124-557-58 | Sequence 58, App1 |
| 25 | 192 | 7.5 | 1140 | 12 | US-10-124-557-104 | Sequence 104, App1 |
| 26 | 192 | 7.5 | 1270 | 12 | US-10-124-557-44 | Sequence 44, App1 |
| 27 | 192 | 7.5 | 1311 | 12 | US-10-124-557-42 | Sequence 42, App1 |
| 28 | 192 | 7.5 | 1313 | 12 | US-10-124-557-142 | Sequence 142, App1 |
| 29 | 192 | 7.5 | 1314 | 12 | US-10-124-557-50 | Sequence 50, App1 |
| 30 | 192 | 7.5 | 1320 | 12 | US-10-124-557-46 | Sequence 46, App1 |
| 31 | 192 | 7.5 | 1320 | 12 | US-10-124-557-60 | Sequence 40, App1 |
| 32 | 192 | 7.5 | 1354 | 12 | US-10-124-557-48 | Sequence 48, App1 |
| 33 | 192 | 7.5 | 1361 | 12 | US-10-124-557-40 | Sequence 40, App1 |
| 34 | 192 | 7.5 | 1363 | 12 | US-10-124-557-52 | Sequence 52, App1 |
| 35 | 192 | 7.5 | 1404 | 12 | US-10-124-557-2 | Sequence 2, App1 |
| 36 | 192 | 7.5 | 1404 | 12 | US-10-124-557-62 | Sequence 62, App1 |
| 37 | 191 | 7.5 | 386 | 10 | US-09-864-761-35720 | Sequence 35720, A |
| 38 | 188.5 | 7.4 | 771 | 9 | US-10-007-270-28 | Sequence 28, App1 |
| 39 | 188.5 | 7.4 | 797 | 9 | US-10-007-270-2 | Sequence 2, App1 |
| 40 | 187.5 | 7.3 | 719 | 9 | US-10-007-270-4 | Sequence 4, App1 |
| 41 | 181.5 | 7.1 | 5877 | 9 | US-10-142-515-11 | Sequence 11, App1 |
| 42 | 180 | 7.0 | 699 | 9 | US-10-121-988-143 | Sequence 143, App1 |
| 43 | 179 | 7.0 | 596 | 9 | US-10-063-547-100 | Sequence 100, App1 |
| 44 | 179 | 7.0 | 596 | 9 | US-10-174-590-310 | Sequence 310, App1 |
| 45 | 179 | 7.0 | 596 | 9 | US-10-176-758-310 | Sequence 310, App1 |

ALIGNMENTS

RESULT 1
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLON, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUPFE, DONALD
; APPLICANT: PANICALI, DENNIS
; TITLE OF INVENTION: TUMOR-ASSOCIATED ANTIGEN
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20
Query Match 89.1% Score 2282.5 DB 9: Length 508;
Best Local Similarity 89.2% Pred. No. 5.4e-105;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps 3;
QY 1 MPTQSGPFFLLLT/LTVLTV.....TVYSGHASTPGEKETSATRRSSVPSSTK 51
DB 1 MPTQSGPFFLLLT/LTVLTV.....TVYSGHASTPGEKETSATRRSSVPSSTK 60
QY 52 NAVMTSVLSHSGSGSSTGQGVTLAPATEPPASGAATWGDVTSVPVRRPLSGT 111
DB 61 NAVMTSVLSHSGSGSSTGQGVTLAPATEPPASGAATWGDVTSVPVRRPLSGT 116
QY 112 TPRAHVTSAPNKPAPGASTAPAPAGVTSABDTR-----PPGSTAPAAHGVTSAPDTR 165

| | | | |
|----|-----|---|-----|
| Db | 119 | APPAGVTSADPTRPAPGSTAPPAHVTSAPDTRPAGSTAPPAHVTSADPTR | 178 |
| QY | 166 | PAPGSTAPPAHVTSAPDNRPALGSTAPPVHNVTSAGSAGSASTLVINGTSARATTE | 225 |
| Db | 179 | PAPGSTAPPAHVTSAPDNRPALGSTAPPVHNVTSAGSAGSASTLVINGTSARATTE | 238 |
| QY | 226 | AASKSTPSPBSHSDPTTLASHSTKDASTHSTVPLTSSMHSSTSPOLSTGSEFFEL | 285 |
| Db | 239 | AASKSTPSPBSHSDPTTLASHSTKDASTHSTVPLTSSMHSSTSPOLSTGSEFFEL | 298 |
| QY | 286 | SHSINLQFNSSTLEDSTDYVOELORDISMFQIYKGGFLGHSNIKRKRPSSVVOGLT | 345 |
| Db | 299 | SHSINLQFNSSTLEDSTDYVOELORDISMFQIYKGGFLGHSNIKRKRPSSVVOGLT | 358 |
| QY | 346 | ARECTINWHDVTOFNQYKTEAASRYNLTLSDVSASHVPPFPAAGSGAGVGNGMILLV | 405 |
| Db | 359 | ARECTINWHDVTOFNQYKTEAASRYNLTLSDVSASHVPPFPAAGSGAGVGNGMILLV | 418 |
| QY | 406 | LVCVLVALAIVYTLALAVCCCKRNKYQGLDIPRADTYHPMSEPTVYTHGRVYPPSSD | 465 |
| Db | 419 | LVCVLVALAIVYTLALAVCCCKRNKYQGLDIPRADTYHPMSEPTVYTHGRVYPPSSD | 478 |
| QY | 466 | RSPTPEKVSAGNGSSLSYTNPAVAATSANI | 495 |
| Db | 479 | RSPTPEKVSAGNGSSLSYTNPAVAATSANI | 508 |

| | | | | |
|----|-----|--|-----|-----|
| Dh | 301 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 164 | 164 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 361 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 420 | 420 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 421 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 480 | 480 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 481 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 540 | 540 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 541 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 600 | 600 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 601 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 660 | 660 |
| Qy | 165 | ----- | 164 | 164 |
| Dh | 661 | APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 720 | 720 |

```

RESULT 2
US-09-996-069-10
; Sequence 10, Application US/09996069
; Publication No. US20030036199A1
GENERAL INFORMATION:
APPLICANT: Bamdad, Cynthia
APPLICANT: Bamdad, R. Shoshana
TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS, AND METHODS FOR TREATMENT OF CANCER
FILE REFERENCE: WO1015/700731
CURRENT APPLICATION NUMBER: US/09/996, 069
CURRENT FILING DATE: 2001-11-27
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

```

| | | | |
|----|-----|---|------|
| Db | 721 | ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 780 |
| Qy | 165 | ----- | 164 |
| Db | 781 | ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 840 |
| Qy | 165 | ----- | 164 |
| Db | 841 | ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 900 |
| Qy | 165 | ----- | 200 |
| Db | 901 | ADTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS | 960 |
| Qy | 201 | ASGASGASASTLVHNGTSSARATTTTPACKSTPSPISHSHTDPTTLLASHSTKTKDAESTHS | 260 |
| Db | 961 | ASGASGASGASTLVHNGTSSARATTTTPACKSTPSPISHSHTDPTTLLASHSTKTKDAESTHS | 1020 |

| | | | | |
|------------------------|-----------------|---|-------------|--------------|
| Query Match Similarity | 83.1% | Score 2130; | DB 9; | Length 1255; |
| Best Local Similarity | 39.0%; | Pred. No. 3,8e-97; | | |
| Matches 489; | Conservative 1; | Mismatches 5; | Indels 760; | Gaps 1; |
| Qy | 1 | MPGQSPFFILLITLVLTVTGSGHASTPGEKEXTAOTORSVSSSTKNAVSMSSV | 60 | |
| | | | | |
| Db | 1 | MPGQSPFFILLITLVLTVTGSGHASTPGEKEXTAOTORSVSSSTKNAVSMSSV | 60 | |
| Qy | 61 | LSHSPGSGSSTTQGDVTLAPATERPASSAATWQODVSVTPTRPAGSTTPPAHGYTS | 120 | |
| | | | | |
| Db | 61 | LSHSPGSGSSTTQGDVTLAPATERPASSAATWQODVSVTPTRPAGSTTPPAHGYTS | 120 | |
| Qy | 121 | APDNKPAAGSTAPPAHGYTSAPDTRPBGSTAPPAHGYTSAPDT | 164 | |
| | | | | |
| Db | 121 | APDNKPAAGSTAPPAHGYTSAPDTRPBGSTAPPAHGYTSAPDT | 164 | |
| Qy | 165 | ----- | ----- | 164 |
| | | | | |
| Db | 181 | ADPTRPAGSTAPPAHGYTSAPDTRPBGSTAPPAHGYTSAPDTRPAGSTAPPAHGYTS | 240 | |
| Qy | 165 | ----- | ----- | 164 |
| | | | | |
| Db | 241 | ADPTRPAGSTAPPAHGYTSAPDTRPBGSTAPPAHGYTSAPDTRPAGSTAPPAHGYTS | 300 | |
| Qy | 165 | ----- | ----- | 164 |

| | | | |
|----------|--|--|------|
| Db | 1021 | SVPEPLTSSNMNSTSPOLSTGVSEFFPLTSHISNIGPSSLEDPSTDDYQELGRISMPLOI | 108 |
| Qy | 321 | YKGGFGLG.SNIFKPPGSAVVOLTLAPREGTINNHVDVETQNOYKTEAASRPNLTISPV | 380 |
| Db | 1081 | YKGGFGLG.SNIFKPPGSAVVOLTLAPREGTINNHVDVETQNOYKTEAASRPNLTISPV | 114 |
| Qy | 381 | VSHPDPPPSASQAGHGMGIALMYLVCVVALATVYTLAAVCCCRPRNYGOLDIFPAR | 440 |
| Db | 1141 | VSHPDPPPSASQAGHGMGIALMYLVCVVALATVYTLAAVCCCRPRNYGOLDIFPAR | 120 |
| Qy | 441 | DTHPPHSEVYTHYTHETGRVYPPSSSTDPRPEIKTISAQNGSSLSSTNPAVAATANL | 495 |
| Db | 1201 | DTHPPHSEVYTHYTHETGRVYPPSSSTDPRPEIKTISAQNGSSLSSTNPAVAATANL | 1255 |
| RESULT 3 | | | |
| | US-09-925-301-861 | | |
| | Sequence 861, Application US/09925301 | | |
| | Patent No. US20020052308A1 | | |
| | GENERAL INFORMATION: | | |
| | APPLICANT: Rosen et al. | | |
| | TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies | | |
| | FILE REFERENCE: PA106 | | |
| | CURRENT APPLICATION NUMBER: US/09/925.301 | | |

CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 861
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: 1181
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 64.0%; Score 1640; DB 10; Length 321;
Best Local Similarity 99.1%; Pred. No. 1e-73;
Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 175 ARGVTSAPDNRPALGSTAPPVHNVTSSAGSAGSASTLVHNGTSARATTPPASKSTPSSI 234
DB 1 ARGVTSAPDNRPALGSTAPPVHNVTSSAGSAGSASTLVHNGTSARATTPPASKSTPSSI 60
QY 235 PSHHSDPTPTLASHSTKTDASTHSTVPPPLTSSNHSTSPOLSTGVSEFFFLSPHISNLOF 294
DB 61 PSHHSDPTPTLASHSTKTDASTHSTVPPPLTSSNHSTSPOLSTGVSEFFFLSPHISNLOF 120
QY 295 NSSLEDPSTDYOELODIEMFLQIYKGGFLQLSNIKRPQSVVVOULTAFREGTINV 354
DB 121 NSSLEDPSTDYOELODIEMFLQIYKGGFLQLSNIKRPQSVVVOULTAFREGTINV 180
QY 355 HDVETQNGYKTEAASRYNTLIDSVSHVFPFPPSAGSAGVGVGNGIALVLVGVVLA 414
DB 181 HDVETQNGYKTEAASRYNTLIDSVSHVFPFPPSAGSAGVGVGNGIALVLVGVVLA 240
QY 415 IYVLIILAACCCRRKRYGOLDIPADDTYHMSSEYPTTHHGKGYVPSPGSTRSEYKVA 474
DB 241 IYVLIILAACCCRRKRYGOLDIPADDTYHMSSEYPTTHHGKGYVPSPGSTRSEYKVA 300
QY 475 GNGGSLSTTPNPAVATSANL 495
DB 301 GNGGSLSTTPNPAVATSANL 321

RESULT 4
US-09-965-131-6
Sequence 6, Application US/09965131
Patent No. US20020160502A1
GENERAL INFORMATION:
APPLICANT: Chung, Maureen A.
APPLICANT: Sharma, Surendra
APPLICANT: Chang, Helena R.
APPLICANT: O'Donnell, Mark A.
TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE
PREVENTION AND TREATMENT OF CANCER
FILE REFERENCE: WI-014CP
CURRENT APPLICATION NUMBER: US/09/965,131
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/235,455
PRIOR FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PaasSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 100
TYPE: PRT
ORGANISM: Homo sapiens
US-09-965-131-6

Query Match 18.1%; Score 465; DB 9; Length 100;
Best Local Similarity 87.9%; Pred. No. 1.2e-16;
Matches 87; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 98 VTSVPVTRPALGSTTPPAHDVTSAPDNKPAGSTAPPAGHGVTSAPDTRPPPGSTAPAAHG 157
DB 2 VTSAPDTRPPAGSTAPPAGHGVTSAPDTRPPAGSTAPPAGHGVTSAPDTRPPAGSTAPPAGH 61
QY 158 VTSAPDTRPPAGSTAPPAGHGVTSAPDNRPALGSTAPPVH 196
DB 62 VTSAPDTRPPAGSTAPPAGHGVTSAPDTRPPAGSTAPPVH 100

RESULT 5
US-10-149-819-7
Sequence 7, Application US/10149819
Publication No. US20030044913A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: YUE, Henry
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: PATTERSON, Chandra
APPLICANT: BAIGER, Mariah R.
APPLICANT: LU, Dying Aina M.
APPLICANT: SHAH, Pulvi
APPLICANT: LAU, Preeti
APPLICANT: AU-YOUNG, Janice
APPLICANT: BURFORD, Neil
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
FILE REFERENCE: PF-0760 PCT
CURRENT APPLICATION NUMBER: US/10/149,819
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
PRIOR FILING DATE: 1999-12-10; 1999-12-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 7
LENGTH: 377
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 1825182CD1
US-10-149-819-7

Query Match 9.4%; Score 240; DB 9; Length 377;
Best Local Similarity 30.5%; Pred. No. 4.4e-05;
Matches 90; Conservative 44; Mismatches 91; Indels 70; Gaps 17;
QY 37 TSATGRSVPSSTKRNVAVMTSSVLSHSPGSGSTTQGDVTLATPAPASGATMGQ 96
DB 118 SASRTITTLFNVY--FLPQNAFTTSSRENVNTSSV--ATLAPVNSPAP-----Q 163
QY 97 DVTSVPTTRPALGSTTPPAHDVTSAPDNKPAPOST--APPAHGVTSAPDTRPPPGSTAPA 154
DB 164 DTTAAP--PTPSATTP-----APSSAPPETTAAP--TPSATTOAPSSAP 209
QY 155 ARGVTSAPDTRPA-----GSTAPPAGHVSAPDNRP-----LGSAPPVHNVTSSAGS 204
DB 210 E--TTAAPPTTPPAAPPSSAPPE--TTAAPPTPSATTPAPAPSSAPP-----ETTAVP 261
QY 205 AGSAGSLVHNGTSARATTT-----PASKSTPSPISPHSDT-----PTTLASH 248
DB 262 PTPSATTTDPSSAPAPPETTAAPPPTPSATTPAPAPSSAPAPQETTAAPPPTTPNSPTTLAD 321
QY 249 STKTDASTHSTVPPPLTSSNHSTSPOLSTG--VSFFPLSFHISNLOFNSILED 300
DB 322 TSRTSAPDTHGTTTSVTTQTTTTKQPTSPAGQNKISRFL--YMKNI--LRIID 373

RESULT 6
US-09-745-008-34
Sequence 34, Application US/09745008
Patent No. US20020137667A1
GENERAL INFORMATION:

```

; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Mircelo A.
; TITLE OF INVENTION: T. Cruz1-Derived Neurotrophic Agents and
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-34

Query Match          9.1%; Score 234; DB 10; Length 1162;
Best Local Similarity 28.7%; Pred. No. 0.00026;
Matches 87; Conservative 34; Mismatches 134; Indels 48; Gaps 11;

QY 16 TVLPTVVGSGH-----ASSTPGKEKTSATQSSVPSSTENKAVSMTSVLSGSHS 65
DB 707 TPSTPVDSSAHGTPSTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSAHG 766
QY 66 PGSSSTTGQODVTLAPATPEPSSAAT-----KGQDVTSVPYTRPALGCTTP-----AHD 117
DB 767 TPSTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSAHG 826
QY 118 VTSAPDNKRAPGSTAP---AHGVTAP-DTRPPPGSTAA---AHGVTAPDTRPAPG 169
DB 827 TPSTPVDSSAHGTPSTPADSSAHSTPTPADSSAHSTPTPADSSAHGTPSTPADSSAHG 886
QY 170 STAPP-----AHGVTAPDNKRALGSTAPVHNVTASGASGASTLV 213
DB 887 TPSTPADSSAHSTPTPADSSAHSTPTPADSSAHGCTPTVD---SSAHGTPSTPADSSA 944
QY 214 HNGTSAPATTPPAKSTPPSPSHS--DTPTTLASHST-KTDASTHST-VPPLTSSN 269
DB 945 HSTSTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSA 1004
QY 270 HST 272
DB 1005 HST 1007

RESULT 7
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Paul
; APPLICANT: Silva, Jelt
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match          9.1%; Score 233.5; DB 10; Length 1367;
Best Local Similarity 28.9%; Pred. No. 0.00032;
Matches 88; Conservative 45; Mismatches 123; Indels 49; Gaps 12;

QY 20 VWTGSGHASTPGKEKTSATQSSVPSSTENKAVSMTSVLSGSHS---PGSSSTT 73
DB 344 VTSSTPVDSSAHGTPSTPADSSAHSTPTPADSSAHSTPTPADSSAHSTPTPADSSA 403
QY 74 QGQDVTLAPATEPAPGSAATWG-ODVTSVPYTRPALGCTTPPAHDVTSAPDNKRAP- 128
DB 404 ESSAPVPTSSSTPSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSS 463
QY 129 ---GSTAPPAHGV---SAPDTRPPGSTAPPAHGVTSAPDTR---PAPGSTAPPAH 176
DB 464 STPSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSS 522
QY 177 GVTAPDNKRALGST---APVHNVTASGSA---SGASITVHNGTSAPATTPPAKS- 229
DB 523 ---SSAPATPSSSTPSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPT 580
QY 230 --TPPSTPSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPT 273
DB 581 VPTSSSTPSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPTSSSAPVPT 640
QY 274 POLST 278
DB 641 PSSST 645

RESULT 8
US-09-996-069-3
; Sequence 3, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIB
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; PRIOR FILING DATE: 2001-11-27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-3

Query Match          9.0%; Score 230; DB 9; Length 54;
Best Local Similarity 95.8%; Pred. No. 2.1e-05;
Matches 46; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 341 VQTLAFREGTINVDVETQPNQYKTEAASRNKLTISDVSYSVHPVPPF 388
DB 1 VQTLAFREGTINVDVETQPNQYKTEAASRNKLTISDVSYSVHPVPPF 48

RESULT 9
US-10-025-380-1065
; Sequence 1065, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshan
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretic, Heather
```

```
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darlick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025.380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1123
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 1065
LENGTH: 957
TYPE: PRT
ORGANISM: Homo sapiens
US-10-025-380-1065
```

```
Query Match 9.0%; Score 230; DB 9; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;
```

```
QY 28 SSTPGKEENSATORSVSPSTEKNVAMTSSVL-----SSH-SPGS-----GSTTQG 75
DB 205 SUTTSGRGESSSTSHSST-THTISAPSTTSALVEPTSYHSSPGSATTHPDSSTTSG 263
QY 76 --QDVTLAAPTEPASGAATWGOVTV-----PYTRPAL-----GSTTPP--AND 117
DB 264 REESTASHNQAOTITVLPARSTSVLGEESTSPISSGMETALPSTTTPGLSRHTK 323
QY 118 VTSAPDNKPAAGSTAPPAAGVTSADTRPPPGSTAPAAAGVTSAPDRPARG----- 169
DB 324 STTPHSPRSAPATTLSPASTTSSGVSESTTSHSRPGSTHTTAFTDSTTTPGLSRHTS 383
QY 170 -----STAPPAGVTSAPDNRPALGSTAPPVHNVTSAGS----- 204
DB 384 HSPGSTDTTLPLASTTSSGVESTTSHSPGSTDTALSPGSTTALSFGQESTTFHSSP 443
QY 205 -----ASG--SASTLVHNGT--SARATTPPAKSTP-----PSIPSHNDTP 242
DB 444 GSTHTTLFPDSTTSSGIVEASTRVHSSGTSPRTTLPASTSPGLQGESTAFTHPASTH 503
QY 243 TTLASHSTKT---DASTHSTVPLTSSNHSTSPOLSTGVFFFLSFHISNLOFNSSLE 299
DB 504 TTPSPATAPAVEESTTYHRS--PSSTPTTHFPASSTSGHSEKSTTFHSSP---DASGT 559
QY 300 DSTSTYVLOLORDISEMFIQIKOGFGLGSNKKFRPGSVVVOGLTAFREGTINVDVET 359
DB 560 TSSAHSTSGRGS-----TTSKRISPGSTEL--TTLGSTTTGGLSEAST 603
QY 360 QP-----NOYKTEASRYNLTIDYVSHVPPPSAAGV-- 396
DB 604 TYSSPSPSTTTLSPASMTSLGVGEESTSRSGSTHVSAPSTTTPGLSEESTTVTS 663
QY 397 --PGMGIALLVLCVLAVALIYVLLALAVCCQRRKNYQGLDIFPARDTYHMESEY- TYH 453
DB 664 SSPG-----STETTVFPSTTTSVGRGEPTTH 691
QY 454 -----THGRVPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
DB 692 SRPASTHTTLFTEDSTTSGLTRESTAPFGSPASTGTGLPATLTADL 738
```

```
RESULT 10
US-09-922-217-1065
```

```
Sequence 1065, Application US/09922217
Patent No. US2002007641A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C13
CURRENT APPLICATION NUMBER: US/09/922.217
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 1065
LENGTH: 957
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-217-1065
```

```
Query Match 9.0%; Score 230; DB 10; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;
```

```
QY 28 SSTPGKEENSATORSVSPSTEKNVAMTSSVL-----SSH-SPGS-----GSTTQG 75
DB 205 SUTTSGRGESSSTSHSST-THTISAPSTTSALVEPTSYHSSPGSATTHPDSSTTSG 263
QY 76 --QDVTLAAPTEPASGAATWGOVTV-----PYTRPAL-----GSTTPP--AND 117
DB 264 REESTASHNQAOTITVLPARSTSVLGEESTSPISSGMETALPSTTTPGLSRHTK 323
QY 118 VTSAPDNKPAAGSTAPPAAGVTSADTRPPPGSTAPAAAGVTSAPDRPARG----- 169
DB 324 STTPHSPRSAPATTLSPASTTSSGVSESTTSHSRPGSTHTTAFTDSTTTPGLSRHTS 383
QY 170 -----STAPPAGVTSAPDNRPALGSTAPPVHNVTSAGS----- 204
DB 384 HSPGSTDTTLPLASTTSSGVESTTSHSPGSTDTALSPGSTTALSFGQESTTFHSSP 443
QY 205 -----ASG--SASTLVHNGT--SARATTPPAKSTP-----PSIPSHNDTP 242
DB 444 GSTHTTLFPDSTTSSGIVEASTRVHSSGTSPRTTLPASTSPGLQGESTAFTHPASTH 503
QY 243 TTLASHSTKT---DASTHSTVPLTSSNHSTSPOLSTGVFFFLSFHISNLOFNSSLE 299
DB 504 TTPSPATAPAVEESTTYHRS--PSSTPTTHFPASSTSGHSEKSTTFHSSP---DASGT 559
QY 300 DSTSTYVLOLORDISEMFIQIKOGFGLGSNKKFRPGSVVVOGLTAFREGTINVDVET 359
DB 560 TSSAHSTSGRGS-----TTSKRISPGSTEL--TTLGSTTTGGLSEAST 603
QY 360 QP-----NOYKTEASRYNLTIDYVSHVPPPSAAGV-- 396
DB 604 TYSSPSPSTTTLSPASMTSLGVGEESTSRSGSTHVSAPSTTTPGLSEESTTVTS 663
QY 397 --PGMGIALLVLCVLAVALIYVLLALAVCCQRRKNYQGLDIFPARDTYHMESEY- TYH 453
DB 664 SSPG-----STETTVFPSTTTSVGRGEPTTH 691
QY 454 -----THGRVPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
DB 692 SRPASTHTTLFTEDSTTSGLTRESTAPFGSPASTGTGLPATLTADL 738
```

```
RESULT 11
US-09-833-263-1065
/ Sequence 1065, Application US/09833263
/ Patent No. US20020110547A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Aijun
/ APPLICANT: Wang, Jonathan D.
/ APPLICANT: Clapper, John A.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Meagher, Madeleine J.
/ TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
/ FILE REFERENCE: 210121.471C12
/ CURRENT APPLICATION NUMBER: US/09/833.263
/ NUMBER OF SEQ ID NOS: 1093
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1065
/ LENGTH: 957
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-833-263-1065

Query Match      9.0%; Score 230; DB 10; Length 957;
Best Local Similarity 22.3%; Pred. No. 0.00033;
Matches 131; Conservative 74; Mismatches 210; Indels 172; Gaps 23;

QY 28 SSTPGEKSTATQSSVSPSTKNAVMTSSVL-----SSH-SRGS-----GSTTQG 75
DB 205 SUTTSRGESTSHSSST-THTTSSAPSTTSMALVEPTSHSGSTATHTPPDSSTTSG 263
QY 76 --QDVTAPATPAPASGAATWGDVTSV-----PYTRPAL-----GSTTPP--AHD 117
DB 264 RSEESTSHSNDQATGTVLPARASTTSLVLSGSTTSPSSGSEMTALPGSTTTTGSLSEK 323
QY 118 VTSAADNKPAPASTAPPAHGVTSAPDTPPPGSTAPAAHGVTSAPDTPRAPG----- 169
DB 324 STTHSSPSPSPATITLSPASTTSSGVSSESTTSHSPGSGTHTATFPDSTTTTGLSHSSTTS 383
QY 170 -----STAPPAHGVTSAPDNPALGSTAPPVNVTSASGS----- 204
DB 384 HSPSGSTDTTLTPASTTSSGVSSESTTSHSSTGSTDALSPGSTTALSPGSESTTFHSSP 443
QY 205 -----ASG--SASTLVHNGT--SARATTTTPASKSTP-----PSIPSHHSDTP 242
DB 444 GSTHTTLFPDSTTSSGIVESTRVHSSTGSPRTTLPASTTSPGLQGESESTAPQTHPASTH 503
QY 243 TTLSHSTKT--DASTHSTVPLTSSNHSTSPQSLATGVSFFPLSPHISMLQPNSSLE 289
DB 504 TTPTSTPTATAPVESESTTVHRS--PSSPTTTPASTTSGHSEKSTTFHSSP--DASGT 559
QY 300 DPTSDYQGLQDISEMFLQYKGGFLGSLNIKRPSSVYVQULTAPREGTINHYDET 359
DB 560 TFSASHTTSGKGS-----TSKISPSGSTET--TTLPGSTTTPLGSEAST 603
QY 360 QF-----NQYTEASGRVYMLTSDVSVSHVPPPSAQSGAGY-- 396
DB 604 TFYSSPSPSTTTLSPASMTSLGVESESTTSSQPSSTHSTVSPASTTTPGLSEESTTVYS 663
QY 397 --PCWGIALALVLCVVALAIVLIALAVQCCKRKQYQGLDIPPADTYHPSSEVP--TYH 453
DB 664 SSPG--STETTVFPRSTTSSVGESEPTTFH 691
QY 454 -----THGRVPPSPSTDRSPYKVSAGNGGSSLSYTNPAVAATSAHL 495
DB 692 SRPASHTTTLFTEDSTYGLTEESTAPPGSPASTOTGPAVLTLTAADL 738

RESULT 12
US-10-025-380-1068
/ Sequence 1068, Application US/10025380
/ Patent No. US20020182191A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jianshun
```

```
APPLICANT: Lodes, Michael J.
APPLICANT: Secret, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yudi
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yael A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darrick
/ TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
/ FILE REFERENCE: 210121.471C14
/ CURRENT APPLICATION NUMBER: US/10/025,380
/ NUMBER OF SEQ ID NOS: 1129
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1068
/ LENGTH: 5179
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match      8.5%; Score 216.5; DB 9; Length 5179;
Best Local Similarity 26.0%; Pred. No. 0.0079;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

QY 2 TPTQSPFLLILLITLVVTSAGHASTPGEKETSATQKSSV-----PSTKNAVMT 57
DB 1478 SPPTTP-----SPPTTTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1530
QY 58 SSVLSHSPSGSS--TQGGDVTLPATPAPASGAATWGDVTSVPVTRPALGSTT-- 112
DB 1531 TTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1586
QY 113 -PPADVTAP--DNKPAPGSTAPPAHGVTSAPDTPR--PPGSTA-----PAA 155
DB 1587 SPPTTTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1646
QY 156 HGVTSAPDTPAPGST--APPAGVTSAP--DNRPALGSTAPPVNVTSASGSASGS 208
DB 1647 TTTTPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 1706
QY 209 ASPLVHNGTSARATTPPASKSTPPSPISPHSDPTTLASHSTKTDAS--STHHSVTPPLT 266
DB 1707 -----SPITTTTTPSSSTTP-----SPPTTMTTPSPPTTPPTTPPTTPPTTPPTTP 1751
QY 267 SSNHSTSPQSLSTGVS 281
DB 1752 TSSPTTTPDLPSPIT 1766

RESULT 13
US-09-922-217-1068
/ Sequence 1068, Application US/09922217
/ Patent No. US20020076414A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jianshun
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Secret, Heather
/ APPLICANT: Benson, Darin R.
/ APPLICANT: Meagher, Madeleine Joy
/ APPLICANT: Stolk, John A.
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Jiang, Yudi
/ APPLICANT: Smith, Carole Lynn
/ APPLICANT: King, Gordon E.
/ APPLICANT: Wang, Aijun
```

Page 7

```

Db      1531 TTPPTTPEPPPTTTPIPTPSTTTTLPTTTPSPPTTT-----TTPPTTTPSPPTTTTP 1586
QY      113 -PPAHDTVSA-P---DNKPAGSADPAAGVTSAAPDTRP--PGGTA-----PAA 155
          ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1587 SPPLTTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTPIITPSTLTLPTTTPSPPT 1646
QY      156 HGVTASADTRPARGST---APRAHGVTSAP----DNPALGSNAPPNNMTSASGASGS 208
          ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1647 TTTTTPPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTTSSPTTTSPTTTMTTTPPTT 1706
QY      209 ASLVNHGTSARATTTTPAKSKTTPPSIPSHSDUTPTLTASTKTDA--STHSVDPDLT 266
          :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1707 -----SPITTTTTSPSTTTP-----SPPTTTTTPSPPTTSPPTTMTTTLPTT 1751
QY      267 SSNSHTSPQLSTGVS 281
          ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1752 TSSPLTTTTPLPSPIT 1766

RESULT 15
US-09-996-069-7
; Sequence 7, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, R. Shoshana
; APPLICANT: Bamdad, Cynthia
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-7

Query Match      8.4%; Score 215; DB 9; Length 45;
Best Local Similarity 95.6%; Pred. No. 9.4e-05;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      350 GTINWHDVETOPNOYKTEAARGYLITSDVSVSHVPFPFSAOGGA 394
          ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
DB      1   GTINWHDVETOPNOYKTEAASPVNLITSDVSVSDVPPFPFSAOGGA 45

```


Page 1

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:10 ; Search time 63.8393 Seconds
(without adjustment)

1033.205 Million cell updates/sec

Title: US-09-658-621B-2

Sequence: 1 MTPGTQSPFFLLLLTLTV.....NGGSSLSYTNPAVAATSANL 495

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_101002:*

| | |
|-----|---|
| 1. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1980.DAT.* |
| 2. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1981.DAT.* |
| 3. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1982.DAT.* |
| 4. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1983.DAT.* |
| 5. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1984.DAT.* |
| 6. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1985.DAT.* |
| 7. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1986.DAT.* |
| 8. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1987.DAT.* |
| 9. | /S1D5/gcgatda/genseeq/genseeqp-emb1/AA1988.DAT.* |
| 10. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1989.DAT.* |
| 11. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1990.DAT.* |
| 12. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1991.DAT.* |
| 13. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1992.DAT.* |
| 14. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1993.DAT.* |
| 15. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1994.DAT.* |
| 16. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1995.DAT.* |
| 17. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1996.DAT.* |
| 18. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1997.DAT.* |
| 19. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA1998.DAT.* |
| 20. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA2000.DAT.* |
| 21. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA2001.DAT.* |
| 22. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA2002.DAT.* |
| 23. | /S1D52/gcgatda/genseeq/genseeqp-emb1/AA2003.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|------------------------|
| 1 | 2562 | 100.0 | 495 | 22 | AAU00539 | Human MUC1 polypeptide |
| 2 | 2413 | 94.2 | 475 | 22 | AAU00573 | Human MUC1 polypeptide |
| 3 | 2413 | 94.2 | 475 | 23 | ABBY77416 | Human MUC1 SE0 IIT |
| 4 | 2391 | 93.3 | 473 | 22 | AAE09508 | Human mucin-1 (MUC1) |
| 5 | 2312 | 90.2 | 455 | 21 | AA71024 | Human Mucin 1 (MUC1) |
| 6 | 2311.5 | 90.2 | 577 | 21 | AA71030 | Ubiquitin-B, collagen |
| 7 | 22295 | 89.6 | 455 | 13 | AAK33973 | Transmembrane form |
| 8 | 2282.5 | 89.1 | 508 | 19 | AAW37223 | Transmembrane protein |
| 9 | 1777 | 69.4 | 348 | 13 | AAK7662 | C-terminal protein |
| 10 | 1671 | 65.2 | 327 | 16 | AAK6298 | Glycoprotein 39 C |

| | | | | | | |
|----|--------|------|------|----|----------|---------------------|
| 1 | 1640 | 64.0 | 3.21 | 21 | ABR43416 | Human cancer associ |
| 2 | 1587 | 61.9 | 3.12 | 23 | HM048410 | Human MUC1R consen |
| 3 | 1352 | 52.8 | 2.87 | 13 | AAK27665 | Secreted form of H |
| 4 | 1273 | 49.7 | 2.73 | 17 | AAK89418 | Mucin-derived prod |
| 5 | 1258.5 | 49.1 | 2.55 | 17 | AAK89419 | Mucin-derived prod |
| 6 | 1174 | 45.8 | 2.82 | 17 | AAK89420 | Mucin-derived prod |
| 7 | 1159.5 | 45.3 | 2.64 | 17 | AAK89421 | Mucin-derived prod |
| 8 | 1091.5 | 42.6 | 2.40 | 17 | AAK89422 | Mucin-derived prod |
| 9 | 1086.5 | 42.4 | 2.39 | 17 | AAK89423 | Mucin-derived prod |
| 10 | 1082.5 | 42.3 | 2.38 | 21 | AAV71028 | UrbNtubulin-E. coli |
| 11 | 1076 | 42.0 | 2.16 | 21 | AAV71022 | Human Mucin 1 (MUC |
| 12 | 1076 | 42.0 | 2.16 | 21 | AAV71027 | Human Mucin 1 (MUC |
| 13 | 854 | 33.3 | 1.73 | 21 | AAV71021 | C-terminal region |
| 14 | 834 | 33.3 | 1.80 | 13 | AAK27664 | N-terminal region |
| 15 | 822 | 33.1 | 1.67 | 13 | AAK27661 | N-terminal region |
| 16 | 823 | 33.1 | 1.67 | 13 | AAK27663 | N-terminal region |
| 17 | 645 | 25.2 | 1.29 | 12 | AAK31110 | Human Polymorphic |
| 18 | 629 | 24.6 | 1.25 | 23 | AAK84809 | Human MUC1R consen |
| 19 | 532.5 | 20.8 | 2.16 | 21 | AAV92665 | MUC1-E. coli |
| 20 | 519 | 20.3 | 2.54 | 23 | AAK60088 | Human melanocyte d |
| 21 | 488 | 19.0 | 2.25 | 21 | AAV71029 | UrbNtubulin-E. coli |
| 22 | 488 | 19.0 | 2.25 | 21 | AAV71023 | Human Mucin 1 (MUC |
| 23 | 482 | 18.8 | 1.05 | 16 | AAK60022 | Mucin peptide p105 |
| 24 | 482 | 18.8 | 1.05 | 19 | AAK72657 | Synthetic MUC1 rep |
| 25 | 462 | 18.1 | 1.00 | 23 | AAK76181 | Synthetic MUC1 rep |
| 26 | 452 | 18.1 | 1.00 | 23 | AAK76181 | Synthetic MUC1 rep |
| 27 | 352 | 13.8 | 0.71 | 22 | AAK60507 | Glycophorin-S-frag |
| 28 | 352 | 13.8 | 0.71 | 22 | AAK60548 | Glycophorin-S-frag |
| 29 | 337 | 13.6 | 1.77 | 22 | AAK28836 | Human Mucin-1 (MUC |
| 30 | 337 | 13.2 | 1.60 | 21 | AAV71020 | Human Mucin 1 (MUC |
| 31 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 32 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 33 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 34 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 35 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 36 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 37 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 38 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 39 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 40 | 337 | 13.2 | 1.60 | 21 | AAV71026 | Human Mucin 1 (MUC |
| 41 | 337 | 13.1 | 1.72 | 21 | AAK62985 | UrbNtubulin-E. coli |
| 42 | 334 | 13.0 | 1.60 | 16 | AAK62987 | Human MUC and huma |
| 43 | 282 | 11.0 | 0.96 | 16 | AAK89426 | Glycophorin-13 N |
| 44 | 266 | 10.4 | 0.96 | 17 | AAK89426 | Mucin-derived prod |
| 45 | 262 | 10.2 | 0.96 | 17 | AAK89426 | Mucin-derived prod |
| 46 | 251.5 | 9.8 | 1.03 | 17 | AAK89427 | Mucin-derived prod |

ALIGNMENTS

RESULT 1

| ID | AAU00539 standard; Protein; 495 AA |
|-----|------------------------------------|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | 10 |
| 11 | 11 |
| 12 | 12 |
| 13 | 13 |
| 14 | 14 |
| 15 | 15 |
| 16 | 16 |
| 17 | 17 |
| 18 | 18 |
| 19 | 19 |
| 20 | 20 |
| 21 | 21 |
| 22 | 22 |
| 23 | 23 |
| 24 | 24 |
| 25 | 25 |
| 26 | 26 |
| 27 | 27 |
| 28 | 28 |
| 29 | 29 |
| 30 | 30 |
| 31 | 31 |
| 32 | 32 |
| 33 | 33 |
| 34 | 34 |
| 35 | 35 |
| 36 | 36 |
| 37 | 37 |
| 38 | 38 |
| 39 | 39 |
| 40 | 40 |
| 41 | 41 |
| 42 | 42 |
| 43 | 43 |
| 44 | 44 |
| 45 | 45 |
| 46 | 46 |
| 47 | 47 |
| 48 | 48 |
| 49 | 49 |
| 50 | 50 |
| 51 | 51 |
| 52 | 52 |
| 53 | 53 |
| 54 | 54 |
| 55 | 55 |
| 56 | 56 |
| 57 | 57 |
| 58 | 58 |
| 59 | 59 |
| 60 | 60 |
| 61 | 61 |
| 62 | 62 |
| 63 | 63 |
| 64 | 64 |
| 65 | 65 |
| 66 | 66 |
| 67 | 67 |
| 68 | 68 |
| 69 | 69 |
| 70 | 70 |
| 71 | 71 |
| 72 | 72 |
| 73 | 73 |
| 74 | 74 |
| 75 | 75 |
| 76 | 76 |
| 77 | 77 |
| 78 | 78 |
| 79 | 79 |
| 80 | 80 |
| 81 | 81 |
| 82 | 82 |
| 83 | 83 |
| 84 | 84 |
| 85 | 85 |
| 86 | 86 |
| 87 | 87 |
| 88 | 88 |
| 89 | 89 |
| 90 | 90 |
| 91 | 91 |
| 92 | 92 |
| 93 | 93 |
| 94 | 94 |
| 95 | 95 |
| 96 | 96 |
| 97 | 97 |
| 98 | 98 |
| 99 | 99 |
| 100 | 100 |

AC AAU00539

DT 12-SEP-2001 (first entry)

Human MUC1 polypeptide.

Human; MUC1; antigenic

KW cancer gene therapy; diagnosis; treatment; inflammatory disorder;

XX

XX
XX
XX

XX
15 435 0001

XX
XX

XX 00 SEP 1960 0005 0001340

PR 10-SEP-1999; 99EP-0402237

[illegible]

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
yy

PI Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM, Koozekanani D, Thomas M.

XX

DR WPI, 2001-235187/24.
 DR N-PSDB; AAS00585.
 XX New antigenic polypeptides of MHC-I protein which activate cytotoxic T
 PT lymphocyte proteins and their analogues, useful for identifying a major
 PT histocompatibility complex class I restricted T cell response and for
 PT diagnosing cancer -
 XX
 XX
 PS Claim 1, Page 68-69; 81pp; English.
 CC The sequence represents a human MUC1 polypeptide. Derivative antigenic
 CC peptides of this protein bind at least one major histocompatibility
 CC complex class I (MHC-I) glycoprotein, which activates cytotoxic T
 CC lymphocytes to induce a protective response against tumours. Diagnosis of
 CC cancer involves determining the presence or absence in a host cell of MHC
 CC class I restricted T cell response to a MUC1 derivative, where the
 CC presence of the MHC class I restricted T cell response indicates that the
 CC host has cancer. Measurement of the level of MHC class I restricted T
 CC cell response is also useful to monitor the severity of cancer, a larger
 CC response indicating a more severe cancer. MUC1 derivatives are useful in
 CC cancer therapy and to follow MUC1 specific immune responses in patients
 CC during the course of disease and/or treatment. MUC1 DNA is useful in
 CC cancer gene therapy, vaccination and diagnosis. Compositions of the
 CC sequences are used in vaccines and treatments against cancer or diseases
 CC caused by an immune response, such as an inflammatory disorder, organ
 CC transplant rejection or graft versus host disease.
 CC
 XX Sequence 495 AA:
 SQ
 Query Match 100.0%; Score 2562; DB 22; Length 495;
 Best Local Similarity 100.0%; Pred. No. 5,8e-152;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGKEKETSATORSVPSSTENAVMTSSV 60
 DB 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGKEKETSATORSVPSSTENAVMTSSV 60
 QY LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 DB LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 QY 61 LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 DB 61 LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 QY 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 QY 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 QY 181 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 240
 DB 181 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 240
 QY 241 TPTTLASHTKTDASTHSTVPTLTSSNHTSQSLTGVSPFISITLMOQNSIED 300
 DB 241 TPTTLASHTKTDASTHSTVPTLTSSNHTSQSLTGVSPFISITLMOQNSIED 300
 QY 301 PSTDIYQIHOIISMFLOIYKGGFLGINSIKRPPGVVQTLAFREGTINVDVETQ 360
 DB 301 PSTDIYQIHOIISMFLOIYKGGFLGINSIKRPPGVVQTLAFREGTINVDVETQ 360
 QY 361 PNOYKTBASRYNLITDSVSVHPFPFPAOSGAGVFGKIALVLCVLAIAIYVLLA 420
 DB 361 PNOYKTBASRYNLITDSVSVHPFPFPAOSGAGVFGKIALVLCVLAIAIYVLLA 420
 QY 421 LAVQCGRKNVQGLDIPAPDTHPMSEVPTHTHGRVYVPSSTRSPSYEKVSAKNGSSS 480
 DB 421 LAVQCGRKNVQGLDIPAPDTHPMSEVPTHTHGRVYVPSSTRSPSYEKVSAKNGSSS 480
 QY 481 LSYTNPAVAATSANL 495
 DB 481 LSYTNPAVAATSANL 495

XX AC AAU00573;
 XX 12-SEP-2001 (first entry)
 DT
 XX
 XX Human MUC1 polypeptide variant.
 DE
 XX
 XX Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;
 KW glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;
 KW cancer gene therapy; diagnosis; treatment; inflammatory disorder;
 KW organ transplant rejection; graft versus host disease.
 XX
 XX Homo sapiens.
 OS
 XX WO200118035-A2.
 XX 15-MAR-2001.
 XX 07-SEP-2000; 2000WO-EP08761.
 PF
 XX 08-SEP-1999; 99GB-0021242.
 PR 10-SEP-1999; 99EP-0402237.
 PR 03-MAR-2000; 2000US-0187215.
 XX
 PA (TRGE) TRANSGENE SA.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
 XX
 XX Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM,
 PI Acres B, Thomas M;
 PI
 XX WPI; 2001-235187/24.
 DR
 XX
 PT New antigenic polypeptides of MHC-I protein which activate cytotoxic T
 PT lymphocyte proteins and their analogues, useful for identifying a major
 PT histocompatibility complex class I restricted T cell response and for
 PT diagnosing cancer -
 XX
 XX
 PS Disclosure; Fig 12; 81pp; English.
 XX
 XX The sequence represents a human MUC1 polypeptide variant. Derivative
 CC antigenic peptides of MUC1 protein bind at least one major
 CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
 CC cytotoxic T lymphocytes to induce a protective response against tumours.
 CC Diagnosis of cancer involves determining the presence or absence in a
 CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
 CC where the presence of the MHC class I restricted T cell response
 CC indicates that the host has cancer. Measurement of the level of MHC class
 CC I restricted T cell response is also useful to monitor the severity of
 CC cancer, a larger response indicating a more severe cancer. MUC1
 CC derivatives are useful in cancer therapy and to follow MUC1 specific
 CC immune responses in patients during the course of disease and/or
 CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
 CC diagnosis. Compositions of the sequences are used in vaccines and
 CC treatments against cancer or diseases caused by an immune response, such
 CC as an inflammatory disorder, organ transplant rejection or graft versus
 CC host disease.
 CC
 XX Sequence 475 AA:
 SQ
 Query Match 94.2%; Score 2413; DB 22; Length 475;
 Best Local Similarity 95.6%; Pred. No. 1.1e-142;
 Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;
 QY 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGKEKETSATORSVPSSTENAVMTSSV 60
 DB 1 MTPQTOSPPFLILLITVLTVTGSGHASTPGKEKETSATORSVPSSTENAVMTSSV 60
 QY 61 LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 DB 61 LLSHSPGSGSSTTGGQDVTLPATBPASGSAATWGQDVTSPVTPRPAAGSTTPPADVTS 120
 QY 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180
 DB 121 APDKPAPGSTAPPAHGVTSAPTRPPGSTAPPAHGVTSAPTRPPAGSTAPPAAGVTS 180

```

Db 121 APDNKPAFGSTAP-----AHGVTSA PDTRPAFGSTAPPAHGVT 160
Oy 181 APDNRPALGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 240
Db 161 APDRRPAFGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 220
Oy 241 TPTTLASHSTKTDASTHSTVPPPLTSSNHSSTPDLSTGVFPFLSHSNLOFNSSLED 300
Db 221 TPTTLASHSTKTDASTHSTVPPPLTSSNHSSTPDLSTGVFPFLSHSNLOFNSSLED 280
Oy 301 PSTDYQELORDISEMFLQIYKOGFGLSNIKFRPGSVVQULTLAFREGTINVDVETQ 360
Db 281 PSTDYQELORDISEMFLQIYKOGFGLSNIKFRPGSVVQULTLAFREGTINVDVETQ 340
Oy 361 FNQKTEASRNLTISDVSVSHVPPFPAAGAGVPGMGIALLVLCVLAIVYLIA 420
Db 341 FNQKTEASRNLTISDVSVSHVPPFPAAGAGVPGMGIALLVLCVLAIVYLIA 400
Oy 421 LAVQCRRKNYGQLDIFPARDTYHPMSSEYPTHTGRVYPPSSSTRSPYEKVSAGNGSS 480
Db 401 LAVQCRRKNYGQLDIFPARDTYHPMSSEYPTHTGRVYPPSSSTRSPYEKVSAGNGSS 460
Oy 481 LSTYNPAVAATSANL 495
Db 461 LSTYNPAVAATSANL 475

RESULT 3
ABR77476
ID ABR77476 standard; Protein: 475 AA.
AC ABR77476;
XX
XX 22-JUL-2002 (first entry)
DT
XX
DE Human MUC1 SEQ ID NO 3.
XX
XX Human; mucin 1; MUC1; transmembrane protein; SNP; cancer; cytostatic;
XX single nucleotide polymorphism; haplotyping; genotyping; drug;
XX antitumor;
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX Misc-difference 337 /note= "SNP allelic variation results in Val substituted
XX by Met"
XX
XX WO200226765-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001MO-US30151.
XX
XX 28-SEP-2000; 2000US-236113P.
XX
XX (GENA-) GENA155ANCE PHARM INC.
XX
XX Chew A, Koshiy B;
XX
XX WPI; 2002-402045/43.
XX
XX N-PSDB; ABL60158, ABL60159.
XX
XX New genetic variants of mucin 1, Transmembrane gene, useful in studying
XX expression and function of protein encoded by the gene and for
XX screening drugs to treat diseases e.g. cancer -
XX
XX Claim 27; Fig 3; 75pp; English.
XX
XX The invention relates to a polynucleotide (ABL60158, ABL60159) encoding
XX mucin 1/MUC1 (ABR77476), Transmembrane isogene. The invention describes
XX novel genetic variants of the MUC1 gene. The invention is useful for
XX haplotyping/genotyping the MUC1 gene in an individual and identifying an
XX

```

```

CC association between a trait and at least one of the haplotypes or
CC haplotype pairs of MUC1 gene. MUC1 is useful for studying the expression
CC and function of MUC1 and expressing MUC1 protein for use in screening
CC for candidate drugs to treat diseases related to MUC1 activity and in
CC studying the effect of the variation on the biological activity of MUC1
CC as well as on the binding affinity of candidate drugs targeting MUC1 for
CC the treatment of e.g. cancer. MUC1 is further used by the pharmaceutical
CC research scientist to validate MUC1 as a candidate target for and in
CC design of clinical trials of candidate drugs for treating a specific
CC condition, drug or disease predicted to be associated with MUC1 activity.
CC MUC1 antibodies are useful in a variety of diagnostic and prognostic
CC formats and therapeutic methods.
CC
XX
SQ Sequence 475 AA;
Query Match 94.2%; Score 2413; DB 23; Length 475;
Best Local Similarity 95.6%; Pred. No. 1,1e-142;
Matches 473; Conservative 0; Mismatches 2; Indels 20; Gaps 1;
Oy 1 MTPGCGSPFLLILITVLTVMGSGHASTPGGKETSATQASVSSTREKAVEMTSV 60
Db 1 MTPGCGSPFLLILITVLTVMGSGHASTPGGKETSATQASVSSTREKAVEMTSV 60
Oy 61 LSSHSPGSSSTTQGDVTLAPATEPAGSAAATWGDVTSVPTRPALCSTTPPAHDVTS 120
Db 61 LSSHSPGSSSTTQGDVTLAPATEPAGSAAATWGDVTSVPTRPALCSTTPPAHDVTS 120
Oy 121 APDNKPAFGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 160
Db 121 APDNKPAFGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 160
Oy 181 APDRRPAFGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 240
Db 161 APDRRPAFGSTAPPVHNVTASAGSAGSASTLVHNGTSARATTPPAKSTPPSPISHSND 220
Oy 241 TPTTLASHSTKTDASTHSTVPPPLTSSNHSSTPDLSTGVFPFLSHSNLOFNSSLED 300
Db 221 TPTTLASHSTKTDASTHSTVPPPLTSSNHSSTPDLSTGVFPFLSHSNLOFNSSLED 280
Oy 301 PSTDYQELORDISEMFLQIYKOGFGLSNIKFRPGSVVQULTLAFREGTINVDVETQ 360
Db 281 PSTDYQELORDISEMFLQIYKOGFGLSNIKFRPGSVVQULTLAFREGTINVDVETQ 340
Oy 361 FNQKTEASRNLTISDVSVSHVPPFPAAGAGVPGMGIALLVLCVLAIVYLIA 420
Db 341 FNQKTEASRNLTISDVSVSHVPPFPAAGAGVPGMGIALLVLCVLAIVYLIA 400
Oy 421 LAVQCRRKNYGQLDIFPARDTYHPMSSEYPTHTGRVYPPSSSTRSPYEKVSAGNGSS 480
Db 401 LAVQCRRKNYGQLDIFPARDTYHPMSSEYPTHTGRVYPPSSSTRSPYEKVSAGNGSS 460
Oy 481 LSTYNPAVAATSANL 495
Db 461 LSTYNPAVAATSANL 475

RESULT 4
AAB09508
ID AAB09508 standard; Protein: 473 AA.
AC AAB09508;
XX
XX 19-NOV-2001 (first entry)
DT
XX
XX Human mucin-1 (MUC-1) protein.
XX
XX Human milk fat globule membrane antigen; HMGf; immunostimulant;
XX cytostatic; cell mediated immune response; carcinoma; adenocarcinoma;
XX breast cancer; dendritic cell; vaccine; gene therapy; mucin-1; MUC-1;
XX immunogen.
XX
XX Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal_peptide
FT Protein 22..473
FT /label= Mature_MUC-1_protein
FT Region 132..170
FT /label= VNTN_region
FT Domain 374..401
FT /label= Transmembrane_domain
XX MO200157068-A1.
XX
XX PD 09-AUG-2001.
XX
XX PF 01-FEB-2001; 2001WO-AU00090.
XX
XX PR 01-FEB-2000; 2000AU-0005369.
XX
XX PR 14-JUN-2000; 2000US-0593870.
XX
XX PA (AUST-) AUSTIN RES INST.
XX
XX PI McKenzie IFC, Pletersz GA, Apostolopoulos V;
XX WPI; 2001-541537/60.
XX
XX PT Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a
XX PT an epitope of the non-VNTR, non-leader region of a mucin -
XX
XX PS Disclosure; Fig 1; 84pp; English.
XX
XX CC The patent discloses peptide or polypeptides capable of eliciting
XX CC an immune response, comprising an amino acid sequence corresponding
XX CC to an epitope of the non-central portion of varying numbers of an
XX CC amino acid motif (VNTR), non-leader region of a mucin. The peptides
XX CC of the invention, fusion proteins comprising the peptide and conjugate
XX CC compounds with carbohydrate polymers are used to induce a cell mediated
XX CC immune response against mucin in the prevention or treatment of
XX CC carcinoma, preferably adenocarcinoma, most preferably breast cancer.
XX CC They are also used to pulse dendritic cell for in vivo transfer and
XX CC use as a vaccine. They are also used in gene therapy. The present
XX CC protein sequence is human milk fat globule membrane antigen (HMFG),
XX CC mucin-1 (MUC-1).
XX
XX SQ Sequence 473 AA;
XX
Query Match 93.3%; Score 2391; DB 22; Length 473;
Best Local Similarity 95.2%; Pred. No. 2,5e-141;
Matches 471; Conservative 0; Mismatches 2; Indels 22; Gaps 2;

```

```

OY 361 FNOYTEAASRYNLTISDVSVHPFPSPASQAGVPGKGTALVLYCVYALAYLYLA 420
OY 339 FNOYTEAASRYNLTISDVSDVDFPSPASQAGVPGKGTALVLYCVYALAYLYLA 398
OY 421 LAVCGCRKXNYGQULDFPAPDTHHMSRPYTHGRVYPPSTDRSPYKSNAGSGSS 480
OY 399 LNVCGCRKXNYGQULDFPAPDTHHMSRPYTHGRVYPPSTDRSPYKSNAGSGSS 458
OY 481 LSTYTPAYATSNL 495
OY 459 LSTYTPAYATSNL 473
XX
XX RESULT 5
XX ID AAY1024
XX AAY1024 standard; Protein; 455 AA.
XX
XX AC AAY1024;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Human Mucin 1 (MUC-1) protein fragment #5.
XX
XX KW Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector;
XX KW anti-tumour; therapy; immune response; cytostatic; vaccine.
XX
XX OS Homo sapiens.
XX
XX PN WO200025827-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 18-OCT-1999; 99WO-EP07874.
XX
XX PR 30-OCT-1998; 98IT-MI02330.
XX
XX PA (MENA) MENARINI RICERCHE SPA.
XX
XX PI Parente D, Di Massimo AM, De Santis R;
XX
XX DR WPI; 2000-365410/31.
XX
XX DR N-PSDB; AAD00388.
XX
XX PT Composition containing one or more DNA molecules encoding fragments of
XX PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
XX PT anti-tumor therapy -
XX
XX PS Claim 16; Fig 5; 56pp; English.
XX
XX CC The present sequence is a fragment of human Mucin 1 (MUC-1), an
XX CC antigenic protein overexpressed in tumour cells. The sequence was
XX CC obtained from B120 tumour cells. The corresponding DNA sequence was
XX CC cloned into a pMRS30 expression vector and used in pharmaceutical
XX CC composition e.g. vaccine for inducing an antigen-specific anti-tumour
XX CC immune response. Composition containing this DNA molecule is useful in
XX CC anti-tumour therapy of patients affected with tumours characterised
XX CC by high MUC-1 expression.
XX
XX SQ Sequence 455 AA;
XX
Query Match 90.2%; Score 2312; DB 21; Length 455;
Best Local Similarity 95.2%; Pred. No. 2e-136;
Matches 452; Conservative 1; Mismatches 2; Indels 20; Gaps 1;

```

```

Db 115 -----AAGTSAADTNPALGSAAPPAAGVSAEDNRKPLASTAPPAHVTS 160
Qy 201 AGSASGASATLVNNGTSARATTPASSTPPETPEHSDPTPTTLASHSTKTDASTHS 260
Db 161 AGSASGASATLVNNGTSARATTPASSTPPETPEHSDPTPTTLASHSTKTDASTHS 220
Qy 261 TVPPTSSNHSSTPQSTGVSPFPLSPHISNLPNSLEDPTDYQELORDISEMPLQI 320
Db 221 TVPPTSSNHSSTPQSTGVSPFPLSPHISNLPNSLEDPTDYQELORDISEMPLQI 280
Qy 321 YKGGFPLGSLNIRKPRGSVVQTLAFREGTINVDVETQFNQKTEASRYNLTSVS 380
Db 281 YKGGFPLGSLNIRKPRGSVVQTLAFREGTINVDVETQFNQKTEASRYNLTSVS 340
Qy 381 VSHVFPFSSAGSAGVPGMGIALVLVLCVLAIAIYLLALAVCCRRKXVQGLDIPAR 440
Db 341 VSDVFPFSSAGSAGVPGMGIALVLVLCVLAIAIYLLALAVCCRRKXVQGLDIPAR 400
Qy 441 DTYHMSSEYPTTHGRVYPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 495
Db 401 DTYHMSSEYPTTHGRVYPPSTDRSPYEKVSAGNGSSLSYTNPAVAATSANL 455

RESULT 6
AAy71030
ID AAy71030 standard; Protein; 577 AA.
AC AAy71030;
XX
DT 29-AUG-2000 (first entry)
DE Ubiqutin-B. coli lacti-human Mucin 1 fusion protein #5.
XX
KM Ubiqutin; Lacti, beta-galactosidase; fusion protein;
human; Mucin 1; MUC-1; tumour; pMR330 expression vector;
XX anti-tumour; therapy; immune response; cytostatic; vaccine.
OS Chimeric - Homo sapiens.
XX Chimeric - Escherichia coli.
OS
FH Key Location/Qualifiers
FT 1..123 /label= Ubiacti protein
FT /note= "contains ubiqutin sequence fused to a
FT portion of B. coli Lacti"
FT Region 124..577
FT /label= Human_MUC-1_fragment
XX
PN WC0200025827-A2.
XX
FD 11-MAY-2000.
XX
PF 18-OCT-1999; 99MO-BP07874.
XX
PR 30-OCT-1999; 98IT-MI02330.
XX
PA (MENA ) MENARINI RICEKHE SPA.
PI Parente D, Di Massimo AM, De Santis R;
XX
DR WPI; 2000-365410/31.
XX N-PSDB; AAD00394.
XX
OS Composition containing one or more DNA molecules encoding fragments of
FT a Mucin 1 (MUC-1) protein overexpressed in tumor cells; useful in
FT anti-tumor therapy -
XX
PS Claim 18; Fig 11; 56pp; English.
XX
CC The present sequence is a fusion protein consisting
of human Mucin 1 (MUC-1) fragment fused to Ubiacti sequence at the
N-terminus. The Ubiacti sequence consists of ubiqutin from MCF7 cell

```

```

CC line and a portion of B. coli beta-galactosidase (lacti). MUC-1 is an
CC antigenic protein overexpressed in tumour cells. The corresponding
CC DNA sequence is cloned into a pMR330 expression vector and used in
CC pharmaceutical composition e.g. vaccine for inducing an antigen-specific
CC anti-tumour immune response. Composition containing this DNA molecule
CC is useful in anti-tumour therapy of patients affected with tumours
CC characterised by high MUC-1 expression.
SQ Sequence 577 AA;
Query Match 90.2%; Score 2311.5; DB 21; Length 577;
Best Local Similarity 91.4%; Pred. No. 2.9e-136;
Matches 459; Conservative 2; Mismatches 14; Indels 27; Gaps 3;
Qy 1 MTEGTO--SP-----FFLLLLLTLLTVVYSGHASSTPGGEKETSATQSSVPSSTEKXA 53
Db 96 LAENTQTASFPALADSLMQLARQVSRGSGHASTPGGEKETSATQSSVPSSTEKXA 155
Qy 54 VSMTSSVLSHSPGSGSSTTQGDVTLAPATEPAGSAAFWGQDVTSVPVTRPALGSTTP 113
Db 156 VSMTSSVLSHSPGSGSSTTQGDVTLAPATEPAGSAAFWGQDVTSVPVTRPALGSTTP 215
Qy 114 PAADVTSAPDNKPAAGSTAPPAHGYTSAPDTRPPGSGTAPAAHGYTSAPDTRPPAGSTAP 173
Db 216 PAADVTSAPDNKPAAGSTAP-----AHGYTSAPDTRPPAGSTAP 255
Qy 174 PAHGYTSAPDNKPAAGSTAPPAHGYTSASGASGASATLVNNGTSARATTPPAKSTPPS 233
Db 256 PAHGYTSAPDNKPAAGSTAPPAHGYTSASGASGASATLVNNGTSARATTPPAKSTPPS 315
Qy 234 IPSHSDPTPTTLASHSTKTDASTHSTVPPLTSSNNHSTSPLSTGVSPFFLSFHSNLQ 293
Db 316 IPSHSDPTPTTLASHSTKTDASTHSTVPPLTSSNNHSTSPLSTGVSPFFLSFHSNLQ 375
Qy 294 FNSSLDEPTDYQELORDISEMPLQIYKGGFGLGSLNIRKPRGSVVQTLAFREGTIN 353
Db 376 FNSSLDEPTDYQELORDISEMPLQIYKGGFGLGSLNIRKPRGSVVQTLAFREGTIN 435
Qy 354 VHVVEFPNQKTEASRYNLTSIVSVSHVFPFSSAGSAGVPGMGIALVLVLCVLA 413
Db 436 VHVVEFPNQKTEASRYNLTSIVSVSDVFPFSSAGSAGVPGMGIALVLVLCVLA 495
Qy 414 AIYVLLALAVCCRRKXVQGLDIPARDTHPMSEYPTTHGRVYPPSTDRSPYEKVS 473
Db 496 AIYVLLALAVCCRRKXVQGLDIPARDTHPMSEYPTTHGRVYPPSTDRSPYEKVS 555
Qy 474 AGNGSSLSYTNPAVAATSANL 495
Db 556 AGNGSSLSYTNPAVAATSANL 577

RESULT 7
AAK23973
ID AAK23973 standard; Protein; 455 AA.
AC AAK23973;
XX
DT 06-NOV-1992 (first entry)
DE Transmembrane form of H23-ETA antigen.
XX
KM ETA-T; human epithelial antigen; Monoclonal antibody H23; vaccine;
KW malignant tumour; breast cancer; tandem repeat.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..21 /label= signal
FT Protein 22..455 /label= ETA-T
FT Misc-difference 134 /label= Pro, Ala

```

| | |
|---------------------------|--|
| Pt | /note= "natural polymorphism" |
| Ft | 144 /label= Thr, Asn |
| Misc-difference | /note= "natural polymorphism" |
| Pt | 147 /label= Pro, Ala |
| Ft | /note= "natural polymorphism" |
| Xx | MO9207000-A. |
| Pd | 30-APR-1992. |
| Bf | 23-OCT-1991; 91WO-FR00835. |
| Kr | 23-OCT-1990; 90FR-0013101. |
| Ra | (TRGE) TRANSGENE SA. |
| Xk | Chambon P, Hareuveni M, Kiény MF, Lathe R, |
| Dl | WPI, 1992-167097/20. |
| Xc | N-PSTDB; AAQ29276. |
| Pt | Compns. contg. polypeptide antigens recognised by antibody H23 - |
| Ft | for treatment of mammary tumours, also for pox virus compns. for |
| Pt | use in vaccines |
| Xx | Claim 3; Page 15-18; 29pp; French. |
| Cc | The tumour antigen recognised by antibody H23 is aberrantly |
| Cc | expressed in epithelial cells from cancerous mammary tissue in about |
| Cc | 90 per cent of breast cancer cases; in a normal individual |
| Cc | expression is negligible. The antigen exists in two forms: |
| Cc | transmembrane (EYF-R) and secreted (EYF-S). Both forms show a high |
| Cc | degree of polymorphism. A 20 amino acid subunit in EYA can be |
| Cc | randomly repeated up to 80 times. From one subunit to the next, 1 to |
| Cc | 3 amino acids can differ. DNA coding for immunogenic fragments of |
| Cc | EYA can be inserted into e.g. vaccinia viruses for treatment of |
| Cc | mammary tumours. |
| Cc | See also AAQ24678-QZ4681, AAQ29277 and AAR23974-R23981. |
| Sq | Sequence 455 AA; |
| Query Match | 89.6%; Score 2295; DB 13; Length 455; |
| Best Local Similarity | 91.3%; Fred. No. 2.3e-135; |
| Matches 452; Conservative | 0; Mismatches 3; Indels 40; Gaps |
| Oy | 1 MPFGTQSPPFLILLLTLVTVTSGSASHSTPGKEKETSATQSSVSPTSKNAVSMTSSV |
| Db | 1 MPFGTQSPPFLILLLTLVLTVTSGSHASTPGEKETSAQSSSVSSTEKNASMTSSV |
| Oy | 61 LSSHSPGSGSSTTGODVTLAATERPASGAATWGODVTSVPVTPALGSTPEPAHDTVS |
| Db | 61 LSSHSPGSGSSTTGODVTLAATERPASGAATWGODVTSVPVTPALGSTPEPAHDTVS |
| Oy | 121 ASDNRPAGSTAEPAHGVTSADPTRRPPGSTAPAHAHVTSADPTRRAPGASTAAPAHGVTS |
| Db | 121 ASDNRPAGSTAEPAHGVTSADPTRRPPGSTAPAHAHVTSADPTRRAPGASTAAPAHGVTS |
| Oy | 181 APDNKPALGSTAPPVHNVTASAGSAGSASTLVENGTGARATTASPASKTPEPSHSND |
| Db | 145 ----RPXIGSTAPPVHNVTASAGSAGSASTLVNHNGTSADAATTAPSASTPEPSHSND |
| Oy | 241 TETTLASHTKTDASTHSTNVPLITSNSHSTSPOLSVGSPFFLISPHIINLOFNSSLSD |
| Db | 201 TETTLASHTKTDASTHSTNVPLITSNSHSTSPOLSVGSPFFLISPHIINLOFNSSLSD |
| Oy | 301 PSTDYOEGLORDISEMFLDIYIKOGGFGLSNKRPRPSVVQLTAPEEGTIIVHDVEIQ |
| Db | 261 PSTDYOEGLORDISEMFLDIYIKOGGFGLSNKRPRPSVVQLTAPEEGTIIVHDVEIQ |
| Oy | 361 FNYQTAASTRNYLTISDVYSNIHFPPPSAAGSAGTCMGAILLVLCUTVALATYTILA |

```

Db      321  FNKCKEALASRYNLITSDVSIVHPFPSSAGSGAGVPGMGALLVLVCLVLAIVLIA 380
Oy      421  LAVOCGRKNNYQGLDIPADDTYHMSRSPRYTHTRGVYPPSSNDSPREKYSAGNGSS 480
Db      381  LAVOCGRKNNYQGLDIPADDTYHMSRSPRYTHTRGVYPPSSNDSPREKYSAGNGSS 440
Oy      481  LSYTPAYVAATSANL 495
Db      441  LSYTPAYVAATSANL 455

RESULT 8
AAW77233
ID      AAW77233 standard; Protein, 508 AA.
AC      AAW77233;
DT      20-NOV-1998 (first entry)
DE      MiniMUC1 protein.
OS      Mus musculus.
KW      MUC1; recombinant pox virus; cytotoxic T-lymphocyte; immunogen;
        tumour; tumour-associated antigen.
XX      Homo sapiens.
XX      MO9837095-A2.
XX      27-AUG-1998.
XX      24-FEB-1998; 98MO-US03693.
XX      24-FEB-1997; 97US-0038253.
XX      (DAMD) DANA FARBER CANCER INST INC.
XX      (THER-) THERION BIOLOGICS CORP.
XX      (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI      Grizic L, Kantor J, Kufe D, Panicali D, Schlom J;
DR      MPI, 1998-467492/40.
DR      N-PSDB; AAW48329.
PT      New recombinant pox virus for tumour therapy - comprises DNA
PT      encoding an immunogenic mini-MUC1 fragment comprising 5-25 MUC1
PT      tandem repeat units
PS      Example 1; Page 21-22; 42pp; English.
XX      The immunogenic mini-MUC1 fragment was created from MUC1 tandem repeat
CC      units for inclusion in a recombinant pox virus (RPV). The RPV was used
CC      in a pharmaceutical composition also containing an immunomodulator to
CC      generate MUC1 specific cytotoxic T-lymphocytes. The recombinant pox
CC      virus therefore encodes an immunogenic MUC1 fragment that does not
CC      undergo significant genetic deletion, thereby providing an unexpectedly
CC      stable and immunogenic pox virus. They can be used to prevent or treat
CC      tumours expressing MUC1 tumour-associated antigens.
XX      Sequence 508 AA;
SQ      Sequence 508 AA;
        89.1%; Score 2282.5; DB 19; Length 508;
Query Match Best Local Similarity 89.1%; Pred. No. 1.6e-134;
Matches 455; Conservative 6; Mismatches 32; Indels 17; Gaps

Oy      1  MPTGTGSPFLILLITLV-----TIVTGGGSHASTPGGKERTSATQSRSSVSSSTEK 51
Db      1  MPTGTGSPFLILLITLVLTTRTKPAIVVYTGSGHASTPGGKERTSATQSRSSVSSSTEK 60
Oy      52  NNVSVMSVLSHSPGSGSSITGQDYLAAAPLEPSSGAATWQDVTVPVTPPALGST 111
Db      61  NNVSVMS--LISNSGVMKSTAPAPAGVTSAPDTRPAGSTAPAGVTSAPDTRPAREST 118
Oy      112  TTPAHDTSTAPDKPASTAPAGVTSAPDTR-----PPGSTRAPAAHGVTSAVDTR 165

```

```

Db 119 APPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTR 178
Qy 166 PAPGSTAPPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTR 225
Db 179 PAPGSTAPPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTR 238
Qy 226 ASKSTPPSPHSHSDPTTTLASHSTKTDASTHSHSTVPLTSSNMHSTSPOLSTGVSPFEL 285
Db 239 ASKSTPPSPHSHSDPTTTLASHSTKTDASTHSHSTVPLTSSNMHSTSPOLSTGVSPFEL 298
Qy 286 SFHISNLOFSSLEDPSTDYOELOORDISEMFLQIYKQGGFLGSLNFKRPGSVVQVLT 345
Db 299 SFHISNLOFSSLEDPSTDYOELOORDISEMFLQIYKQGGFLGSLNFKRPGSVVQVLT 358
Qy 346 AFRREGTINVDVETQFNQYKTEASRYNLITSDVSYSVHPPEFSAQSGAGVPQWGIALLV 405
Db 359 AFRREGTINVDVETQFNQYKTEASRYNLITSDVSYSVHPPEFSAQSGAGVPQWGIALLV 418
Qy 406 LVCVAVALAIYVLIALAVOCCKRKYGQLDIPFARDTYHPMSERYPTVHTGRVPPSSSTD 465
Db 419 LVCVAVALAIYVLIALAVOCCKRKYGQLDIPFARDTYHPMSERYPTVHTGRVPPSSSTD 478
Qy 466 RSPYEKVGAGSGSSSYTNPAVAATSANL 495
Db 479 RSPYEKVGAGSGSSSYTNPAVAATSANL 508

```

```

RESULT 9
AAR27662
ID AAR27662 standard; Protein; 348 AA.
XX
AC AAR27662;
XX
DT 06-NOV-1992 (first entry)
XX
DE C-terminal region of H23-ETA-T antigen.
XX
KW Transmembrane; human epithelial antigen; Monoclonal antibody H23;
KW vaccine; malignant tumour; breast cancer; tandem repeat.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT Misc-difference 7 /label= Pro, Ala
FT Misc-difference 17 /note= "natural polymorphism"
FT Misc-difference 17 /label= Thr, Asn
FT Misc-difference 20 /note= "natural polymorphism"
FT Misc-difference 20 /label= Pro, Ala
FT Misc-difference 27 /note= "natural polymorphism"
FT Misc-difference 27 /label= Pro, Ala
FT Misc-difference 37 /note= "natural polymorphism"
FT Misc-difference 37 /label= Thr, Asn
FT Misc-difference 40 /note= "natural polymorphism"
FT Misc-difference 40 /label= Pro, Ala
FT Region /note= "natural polymorphism"
FT Region /note= "contains 2 tandem repeats"
FT Region /note= "contains up to 80 copies"
XX
XX WO9207000-A.
XX 30-APR-1992.
XX 23-OCT-1991; 91WO-FR00835.
XX 23-OCT-1990; 90FR-0013101.
XX

```

```

XX (TRGE ) TRANSGENE SA.
PA Chambon P, Hareuveni M, Kieny MP, Lathie R;
PI WPI; 1992-167097/20.
XX N-PSDB; AA024679.
DR
XX Compens. contg. polypeptide antigen recognised by antibody H23 -
PT for treatment of mammary tumours, also for pox virus compens. for
PT use in vaccines
PS Claim 3; Page 15-18; 29pp; French.
XX
CC The tumour antigen recognised by antibody H23 is aberrantly
CC expressed in epithelial cells from cancerous mammary tissue in about
CC 90 per cent of breast cancer cases; in a normal individual
CC expression is negligible. The antigen exists in two forms:
CC transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high
CC degree of polymorphism. A 20 amino acid subunit in ETA can be
CC tandemly repeated up to 80 times. (N.B. two tandem repeats are
CC shown here; the first half of the amino acid sequence, i.e. on
CC the N-terminal side of the repeat region, is given in AAR27661.
CC From one subunit to the next, 1 to 3 amino acids can differ.
CC See also AA024678-Q24681, AA029276-7 and AAR23974-R23981.
XX
SQ Sequence 348 AA;

```

```

Query Match 69.4%; Score 1777; DB 13; Length 348;
Best Local Similarity 98.3%; Pred. No. 3.4e-103;
Matches 342; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 148 PGSTAPPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTR 207
Db 1 PGSTAPPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTRPAHGVTSAPDTR 60
Qy 208 SASTLVNCTSAKATTPPAKSTPPSPISHSHTPTTLASHSTKTDASTHSHSTVPLTSS 267
Db 61 SASTLVNCTSAKATTPPAKSTPPSPISHSHTPTTLASHSTKTDASTHSHSTVPLTSS 120
Qy - 268 SNHSTSPOLSTGVSPFELSFHISNLOFNSLSLEDPSTDYOELOORDISEMFLQIYKQGGFL 327
Db 121 SNHSTSPOLSTGVSPFELSFHISNLOFNSLSLEDPSTDYOELOORDISEMFLQIYKQGGFL 180
Qy 328 GLSNIKFRGSAVVQVLTAFREGTINVDVETQFNQYKTEASRYNLITSDVSYSVHPPE 387
Db 181 GLSNIKFRGSAVVQVLTAFREGTINVDVETQFNQYKTEASRYNLITSDVSYSVHPPE 240
Qy 388 FSAQSGAGVPQWGIALLVLCVAVALAIYVLIALAVOCCKRKYGQLDIPFARDTYHPMS 447
Db 241 FSAQSGAGVPQWGIALLVLCVAVALAIYVLIALAVOCCKRKYGQLDIPFARDTYHPMS 300
Qy 448 EYPTVHTGRVPPSSSTDSPYEKVGAGSGSSSYTNPAVAATSANL 495
Db 301 EYPTVHTGRVPPSSSTDSPYEKVGAGSGSSSYTNPAVAATSANL 348

```

```

RESULT 10
AAR96298
ID AAR96298 standard; Peptide; 327 AA.
XX
AC AAR96298;
XX
DT 26-JUL-1996 (first entry)
XX
DE Glycoprotein 39 C terminal fragment.
XX
KW Glycoprotein 39; gp39; lambda gvi1 cDNA library; gastric cancer;
KW cell line KATO-III; tumour; immune abnormality; marker;
KW inflammatory disease.
XX
OS Homo sapiens.
XX

```

```
PN JP07051065-A.
XX
XX 28-FEB-1995.
XX
XX 21-FEB-1992; 92JP-0035085.
XX
XX 21-FEB-1992; 92JP-0035085.
XX
XX 21-FEB-1992; 92JP-0035085.
XX
XX (NIKO-) NIPPON KOTAI KENKUSHO KK.
XX (UYKA-) UNIV KAGOSHIMA.
XX
XX WPI; 1995-127356/17.
XX
XX N-PSDB; AAT29190.
XX
XX Glyco:protein 39 gene - used in the mass production of glyco:protein
XX 39, for use as tumour and immune abnormality marker
XX
XX Claim 2; Page 9-10; 10pp; Japanese.
XX
XX The sequences given in AAR96297-98 represent portions of glycoprotein
XX 39. This sequence represents the C terminal portion of the glycoprotein
XX 39 gene. The cDNA encoding this sequence was isolated from lambda gt11
XX cDNA library derived from the gastric cancer cell line KATO-III
XX Glycoprotein 39 is expected to be used as a tumour marker, an immune
XX abnormality marker or a marker for various inflammatory diseases.
XX
XX Sequence 327 AA;
XX
XX Query Match 65 28; Score 1671; DB 16; Length 327;
XX Best Local Similarity 99.1%; Pred. No. 1.3e-96;
XX Matches 324; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 169 GSTPPAPGCTAPDNRPALGSTAPVHNVTASGSGASGASTLVNNGTSARATTPPASK 228
XX 1 GSTPPAPGCTAPDNRPALGSTAPVHNVTASGSGASGASTLVNNGTSARATTPPASK 60
XX 229 STPTSTSHSTPTTLASHTKTDASTHSTVPTLTSSNSTSPOLSTGVSPFSTSR 288
XX 61 STPTSTSHSTPTTLASHTKTDASTHSTVPTLTSSNSTSPOLSTGVSPFSTSR 120
XX 289 ISNIGRSSLEPSTDYQELQNDISEMFLQYKGGFLGLSNIKRPGSVVQGLTAAR 348
XX 121 ISNIGRSSLEPSTDYQELQNDISEMFLQYKGGFLGLSNIKRPGSVVQGLTAAR 180
XX 349 EGTINHDVETOPNOYKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALVLCV 408
XX 181 EGTINHDVETOPNOYKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALVLCV 240
XX 409 VLVALAIVYLALAVCCRRKNTGOLDIFPARDTYHMSSEYPTTHGRVYVPSSTRSP 468
XX 241 VLVALAIVYLALAVCCRRKNTGOLDIFPARDTYHMSSEYPTTHGRVYVPSSTRSP 300
XX 469 YEKVSAGNGSSLSYTNPAVAATSAWL 495
XX 301 YEKVSAGNGSSLSYTNPAVAATSAWL 327
XX
XX RESULT 11
XX AAB43416
XX ID AAB43416 standard; Protein; 321 AA.
XX
XX AAB43416;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human cancer associated protein sequence SEQ ID NO:861.
XX
XX Human, cancer associated gene; cancer antigen; detection; cancer;
XX diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antiallergic; antithrombotic; antitumor; antiviral;
XX antiinflammatory; antihypertensive; antidiabetic; antidiabetic;
XX dermatologic; neuroprotective; thrombolytic; coagulant; nociceptive;
XX vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
```

```
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KM neurological disease; drug screening.
XX
XX Homo sapiens.
XX
XX WO20005350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587533/55.
XX N-PSDB; AAC78625.
XX
XX Novel isolated nucleic acid comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 11; Page 1414-1416; 2352pp; English.
XX
XX AAC78607 to AAC78448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antiallergic; antithrombotic; antitumor; antiviral;
XX dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;
XX neurotrophic; vasotrophic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 321 AA;
XX
XX Query Match 64.0%; Score 1640; DB 21; Length 321;
XX Best Local Similarity 99.1%; Pred. No. 1.1e-94;
XX Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 175 AHGVTAPDNRPALGSTAPVHNVTASGSGASGASTLVNNGTSARATTPPASKPTPSI 234
XX 1 AHGVTAPDNRPALGSTAPVHNVTASGSGASGASTLVNNGTSARATTPPASKPTPSI 60
XX 235 PSHHSDPTTLASHTKTDASTHSTVPTLTSSNSTSPOLSTGVSPFSTSR 294
XX 61 PSHHSDPTTLASHTKTDASTHSTVPTLTSSNSTSPOLSTGVSPFSTSR 120
XX 295 NSLSLEPSTDYQELQNDISEMFLQYKGGFLGLSNIKRPGSVVQGLTAARFEGTIN 354
XX 121 NSLSLEPSTDYQELQNDISEMFLQYKGGFLGLSNIKRPGSVVQGLTAARFEGTIN 180
XX 355 HDVETOPNOYKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALVLCVVALA 414
XX 181 HDVETOPNOYKTEAASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALVLCVVALA 240
XX 415 IYVILALAVCCRRKNTGOLDIFPARDTYHMSSEYPTTHGRVYVPSSTRSPYEKVA 474
```


PT for treatment of mammary tumours, also for pox virus compans. for
 use in vaccines

PS Claim 3; Page 19-21; 29pp; French.

CC The tumour antigen recognised by antibody H23 is aberrantly
 CC expressed in epithelial cells from cancerous mammary tissue in about
 CC 90 per cent of breast cancer cases; in a normal individual
 CC expression is negligible. The antigen exists in two forms:
 CC transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high
 CC degree of polymorphism. A 20 amino acid subunit in ETA can be
 CC tandemly repeated up to 80 times. From one subunit to the next, 1 to
 CC 3 amino acids can differ. DNA coding for immunogenic fragments of
 CC ETA can be inserted into e.g. vaccinia viruses for treatment of
 CC mammary tumours.
 CC See also AAQ24678-Q24681, AAQ29276-7 and AAR3974-R23981.

XX Sequence 287 AA;

Query Match 52.8%; Score 1352; DB 13; Length 287;
 Best Local Similarity 86.1%; Pred. No. 8.7e-77;

Matches 272; Conservative 0; Mismatches 4; Indels 40; Gaps 1;

QY 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 DB 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 QY 61 LSHSPGSGSSTTGQDVTLAATPAPASGSAATWGQDVTSPVTPPALGSTTPPAHDVTS 120
 DB 61 LSHSPGSGSSTTGQDVTLAATPAPASGSAATWGQDVTSPVTPPALGSTTPPAHDVTS 120
 QY 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 180
 DB 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 180
 QY 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 240
 DB 145 ---FPXGSAAPVAVNVTSAGSASGASASTLVNNGTSARATTPASCTPPGTPSHSD 200
 QY 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 300
 DB 201 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 260
 QY 301 PSTDYVOELQORDISEM 316
 DB 261 PSTDYVOELQORDISEM 276

RESULT 14

AAR89418 standard; Protein; 273 AA.

XX AAR89418;

XX 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/X.

KM MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 KM MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.

XX Homo sapiens.

XX MO9603502-A2.

XX 08-FEB-1996.

XX 21-JUL-1995; 95MO-IB00627.

XX 26-JUL-1994; 94IL-0110464.

XX (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

XX Wreschner DH;

XX WPI; 1996-117047/12.

XX N-PSDB; AAT10677.

PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer

PS Claim 4; Fig 5A; 79pp; English.

CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obcd. by expression of encoding cDNA
 CC (see AAT10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.

XX Sequence 273 AA;

Query Match 49.7%; Score 1273; DB 17; Length 273;
 Best Local Similarity 54.9%; Pred. No. 6.9e-72;

Matches 272; Conservative 1; Mismatches 1; Indels 222; Gaps 1;

QY 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 DB 1 MTPGQSPFFLLLLTLVTVVSGHASTPGGKETSATORSVPSSTERNVMTSSV 60
 QY 61 LSHSPGSGSSTTGQDVTLAATPAPASGSAATWGQDVTSPVTPPALGSTTPPAHDVTS 120
 DB 61 LSHSPGSGSSTTGQDVTLAATPAPASGSAATWGQDVTSPVTPPALGSTTPPAHDVTS 120
 QY 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 180
 DB 121 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 180
 QY 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 240
 DB 181 APDNKPAAGSTAPPAHGVTSAPDTPPPGSAAPAAHGVTSAPDTPAPGSAAPPAHGVTS 240
 QY 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 300
 DB 241 TPTTLASHSTKTDASTHSTVPPVLTSSNHSSTPOLSTGVSPFPLSPHISNLOPNSLSD 260
 QY 301 PSTDYVOELQORDISEM 316
 DB 261 PSTDYVOELQORDISEM 276
 QY 361 FNOYKTEASRYNLTIISDVSVSHVPFPSSAQSAGAVPGGIALLVLCVLAALAYLLA 420
 DB 139 FNOYKTEASRYNLTIISDVSVSHVPFPSSAQSAGAVPGGIALLVLCVLAALAYLLA 198
 QY 421 LAVCOCRRKNGQLDIFPARDTYHMSSEYPTHTGRYVPPSTDRSRYEKVSAGNGGSS 480
 DB 199 LAVCOCRRKNGQLDIFPARDTYHMSSEYPTHTGRYVPPSTDRSRYEKVSAGNGGSS 258
 QY 481 LSTYTPAVVAATSNL 495
 DB 259 LSTYTPAVVAATSNL 273

RESULT 15

AAR89419 standard; Protein; 282 AA.

XX AAR89419;

XX 24-APR-1996 (first entry)

Thu May 8 16:13:59 2003

us-09-658-621b-2.rag

Page 11

DE Mucin-derived protein MUC1/X/alt.
 XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
 KM MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
 KM receptor; diagnosis; imaging; therapy.
 OS Homo sapiens.
 XX
 PN M09603502-A2.
 XX
 PD 08-FEB-1996.
 XX
 PD 21-JUL-1995; 95MO-IB00627.
 XX
 PF 26-JUL-1994; 94IL-0110464.
 XX
 FR
 XX (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.
 PA
 XX Wreschner DH;
 PI
 XX MPI; 1996-117047/12.
 XX
 DR N-PSDB; AAT10678.
 XX
 PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
 PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
 PT cancer
 PS
 PS Claim 4; Fig 5B; 79pp; English.
 CC
 CC Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
 CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR9418-23) are highly and
 CC differentially expressed in human breast cancer tissue and significantly
 CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
 CC They serve as cell surface receptor molecules participating in signal
 CC transduction. The proteins can be obtd. by expression of encoding cDNA
 CC (see ARI10677-82) in recombinant host cells. They are used in the
 CC treatment of human breast cancer and as diagnostic reagents. Receptor
 CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
 CC isolated.
 CC
 CC Sequence 282 AA;
 CC

[illegible]

Db 139 IMWDEVETQNFQKTEAABRNVLITISVUSVDPPEPPFQAQSGACVAPGWCIALIVVCYL 198
Qy 412 ALATLYTLALAVOCGRKNYGGDIPAPRTYHPMSSEPTHTGRVPSSTDSPEYK 471
Db 199 AALATYTLALAVOCGRKNYGGDIPAPRTYHPMSSEPTHTGRVPSSTDSPEYK 256
Qy 472 VSANGSGSSLYTPNPAVATSNL 495
Db 259 VSANGSGSSLYTPNPAVATSNL 282

Search completed: May 1, 2003, 23:10:26
Job time : 66.8393 secs

Page 1

GenCore version 5.1.4_p5_4578
(c) 1993 - 2003 CompuGen Ltd.

arch time 105.089 Seconds

970.540 Million cell updates/sec

.....NGGSSLSYTNPAVAATSANL 495

BLOSUM62

671580

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.podent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.rvms:*
- 16: sp.bacterioph:*
- 17: sp.archaealp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Length | DB | ID | Description |
|------------|--------|-------|--------|----|--------|---------------------|
| 1 | 1795.5 | 70.1 | 553 | 6 | Q9MGL1 | Q9MGL1 macaca mula |
| 2 | 1234 | 48.2 | 602 | 6 | Q19115 | Q19115 oryctolagus |
| 3 | 1228.5 | 48.0 | 580 | 6 | Q9MML4 | Q9MML4 bos taurus |
| 4 | 1198 | 46.8 | 631 | 11 | Q99K60 | Q99K60 mus musculus |
| 5 | 1120.5 | 43.7 | 554 | 6 | Q95I89 | Q95I89 bos taurus |
| 6 | 720.5 | 28.1 | 153 | 11 | Q60S51 | Q60S51 mesocricetus |
| 7 | 716 | 27.9 | 192 | 6 | Q28078 | Q28078 bos taurus |
| 8 | 706.5 | 27.6 | 193 | 6 | Q28723 | Q28723 oryctolagus |
| 9 | 699.5 | 27.3 | 193 | 11 | Q60408 | Q60408 cavia cutlea |
| 10 | 358 | 14.0 | 74 | 11 | Q35770 | Q35770 rattus norvi |
| 11 | 267.5 | 10.4 | 487 | 2 | Q9ZBK9 | Q9ZBK9 ureaplasma |
| 12 | 265.5 | 10.4 | 1334 | 16 | Q9RKR9 | Q9RKR9 streptococ |
| 13 | 264.5 | 10.3 | 1079 | 5 | Q9N5A7 | Q9N5A7 caenorhabd |
| 14 | 254.5 | 9.9 | 160 | 6 | Q8SQ37 | Q8SQ37 bos taurus |
| 15 | 243.5 | 9.9 | 125 | 5 | Q9BXV7 | Q9BXV7 leishmania |
| 16 | 258.5 | 9.7 | 160 | 6 | Q8SQ36 | Q8SQ36 bos taurus |

| | | | | | | |
|----|-------|-----|------|--------|------------|--------------------|
| 17 | 244.5 | 9.5 | 1349 | 4 | Q8NM04 | Q8Wwq4 homo sapien |
| 18 | 240 | 9.4 | 961 | 3 | Q22223 | emericella |
| 19 | 239 | 9.3 | 160 | 6 | Q8G038 | bos taurus |
| 20 | 237 | 9.3 | 2232 | 5 | P91365 | P91365 caenorhabd |
| 21 | 234.5 | 9.2 | 907 | 12 | Q66537 | hunan heppe |
| 22 | 234 | 9.1 | 377 | 4 | Q8TX57 | homo sapien |
| 23 | 233 | 9.1 | 3178 | 5 | Q969D4 | caenorhabd |
| 24 | 233 | 9.1 | 3570 | 4 | Q99552 | homo sapien |
| 25 | 232.5 | 9.1 | 786 | 5 | Q21027 | caenorhabd |
| 26 | 232 | 9.0 | 528 | 4 | Q56854 | human heppe |
| 27 | 231.5 | 9.0 | 957 | 4 | Q29071 | sus scrofa |
| 28 | 230 | 8.9 | 957 | 4 | Q29070 | homo sapien |
| 29 | 229 | 8.9 | 873 | 5 | Q9Y075 | leishmania |
| 30 | 225.5 | 8.8 | 293 | 5 | Q9Y075 | leishmania |
| 31 | 225 | 8.8 | 279 | 4 | Q42988 | homo sapien |
| 32 | 222 | 8.7 | 534 | 3 | Q94317 | echinosach |
| 33 | 222 | 8.7 | 622 | 4 | Q14681 | homo sapien |
| 34 | 221 | 8.6 | 886 | 12 | Q9P087 | human heppe |
| 35 | 220.5 | 8.6 | 1029 | 4 | Q8TH37 | homo sapien |
| 36 | 220.5 | 8.6 | 1612 | 3 | Q9Y671 | neutrosia |
| 37 | 220 | 8.6 | 1592 | 5 | Q9Y072 | dirosophila |
| 38 | 219.5 | 8.6 | 1375 | 5 | Q76602 | caenorhabd |
| 39 | 219 | 8.5 | 6 | Q29070 | sus scrofa | |
| 40 | 218.5 | 8.5 | 322 | 11 | Q62605 | rattus norv |
| 41 | 217.5 | 8.5 | 783 | 12 | Q91331 | ceratophytic |
| 42 | 217 | 8.5 | 1503 | 5 | Q9NSK0 | caenorhabd |
| 43 | 216.5 | 8.5 | 851 | 12 | Q9Q986 | human heppe |
| 44 | 215.5 | 8.4 | 4776 | 16 | Q97971 | streplococ |
| 45 | 215 | 8.4 | 826 | 12 | Q8WO55 | equine hepp |

ALIGNMENTS

RESULT 1

| | | | | | |
|----|--|--|-------------|-----|--------|
| ID | SVN | NAME | PRELIMINARY | PRT | 553 AA |
| AC | 09M2L1 | | | | |
| AD | 09M2L1 | | | | |
| AE | 01-OCT-2000 (TREMBLER) | 15, Created) | | | |
| AF | 01-OCT-2000 (TREMBLER) | 15, Last sequence update) | | | |
| AG | 01-MAR-2000 (TREMBLER) | 15, Last annotation update) | | | |
| AH | Mucin 1 (Fragment) | | | | |
| AI | Mucina m1a1a (Rhesus macaque). | | | | |
| AJ | Eukaryota; Metazoa; Chordata; Cetartia; Vertebrata; Eutelestomi; | | | | |
| AK | Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; | | | | |
| AL | Cercopitheidae; Macaca. | | | | |
| AM | NCBI_TaxID=9544; | | | | |
| AN | 11) | | | | |
| AO | SEQUENCE FROM N.A. | | | | |
| AP | MEDLINE=20330533; PubMed=10869775; | | | | |
| AQ | Vaughan H.A., Ho D.W.M., Karamkas V., Sandtin M.S., McKenzie I.F.C., | | | | |
| AR | Plesters G.A.; | | | | |
| AS | "The Immune Response of Mice and Cynomolgus Monkeys to Macaque Mucin1- | | | | |
| AT | Mannan." | | | | |
| AV | Vaccine 18:3297-3309(2000). | | | | |
| AW | EMBL; AF176947; AAF82403.1; | | | | |
| AX | InterPro: IPR001064; Crystal1in. | | | | |
| AY | InterPro: IPR000082; SEA_domain. | | | | |
| AZ | Pfam; PF01390; SEA; 1. | | | | |
| BA | SMART; SM00200; SEA; 1. | | | | |
| BB | DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1. | | | | |
| BC | PROSITE; PS50024; SEA; 1. | | | | |
| BD | NON_TER | 1 | | | |
| BE | FT | 1 | | | |
| BF | NON_TER | 553 | | | |
| BG | SEQUENCE | 553 AA; 55778 MW; 6D76DD2EB929318 CRC64; | | | |
| BH | Query Match | 70.1%; Score 1795.5; DB 6; Length 553; | | | |
| BI | Best Local Similarity | 67.4%; Pzed. No. 9.7e-102; | | | |
| BJ | Match 379; Conservative | 13; Mismatch 59; Indels 109; Gaps | | | |
| BK | 21. VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| BZ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CX | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CY | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CA | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CB | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CC | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CD | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CE | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CF | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CG | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CH | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CI | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CJ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CK | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CL | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CM | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CN | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CO | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CP | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CQ | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CR | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CS | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CT | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CU | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CV | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| CW | VTGSGHASTPGGEKETSATONRSVPSSTERNNAVSMTSVLSGSHSPSGSGSTTCOGDVTLL | 80 | | | |
| | | | | | |

| | | | |
|----------|---|---|-----|
| Qy | 81 | ADPTSPAGSAAATGQDVTSVVTPEPALGSTTPPAADVYSA.PDNRPA.PAGSTAP.PHAGTYS | 140 |
| Db | 59 | ALATPEAPAGSATYLTGNHVTSAADTSAAPGSGTGPAGVYTSAPDTSAA.PGSGTGP.PARVYTS | 118 |
| Qy | 141 | ADPTPEPPGSGTAPPAAGVTSAPDTPAPGSGTAPPAHGVYTSAPDNRPA----- | 187 |
| Db | 119 | ADPTSAAGSGTGPAPRVYTSAPDTPSAADTSAAPGSGTGPAPVYTSAPDTSAA.PGSGTGP.PARVYTS | 178 |
| Qy | 188 | ----- | 187 |
| Db | 179 | ADPTSAAGSGTGPAPRVYTSAPDTPSAAPGSGTGPAPRVYTSAPDTSAA.PGSGTGP.PARVYTS | 238 |
| Qy | 188 | -----LGSATPHANTYSSAGSAGSASTLVH | 214 |
| Db | 239 | ADPTSAAGSGTGPAPRVYTSAPDTPSAAPGSGTAPPAHGVYTSAPDTSAA.PGSGTGP.PARVYTS | 298 |
| Qy | 215 | NGTSRAATTPPAKSNPPSPBPHSDPTTLAHSKTKDASTHSGVPRPTSNHSTSP | 274 |
| Db | 299 | STTSRAATTPPAKSNPPSPBPHSDPTTLAHSKTKDASTHSGVPRPT-SNHSTSP | 357 |
| Qy | 275 | QJSTGVSEFFSLSPHISNLOPNSLSDPSTDYQYELORDISEMFLQYKQCGFLGSNTKF | 334 |
| Db | 358 | QJSLGVSFFSLSPHISNLOPNSLSDPSTDYQYELORDISEMFLQYKQCGFLGSNTKF | 417 |
| Qy | 335 | PRGSVYVQULTAPEEGTINADVETOFNOYKTEAASRNLTISDVSYSHVPPPSAOSGA | 394 |
| Db | 418 | PRGSVYVQULTAPEEGTINADVETOFNOYKTEAASRNLTISDVSYSHVPPPSAOSGA | 477 |
| Qy | 395 | GVPGKGIALLVLCVLAVALAIYVLLALAVCCQCKRKYGGOLDIPAROTYH.PMSEPTVYHT | 454 |
| Db | 478 | GVPGKGIALLVLCVLAVALAIYVLLALAVCCQCKRKYGGOLDIPARDAVYH.PMSEPTVYHT | 537 |
| Qy | 455 | HGRVYPPSGTDRSEYE | 470 |
| Db | 538 | HGRVYPPAGGTNRSEYE | 553 |
| RESULT 2 | | | |
| OL9115 | | PRELIMINARY; PRT; 602 AA. | |
| ID | OL9115 | | |
| AC | OL9115 | | |
| DT | 01-JAN-1998 | (TREMblrel. 05, Created) | |
| DT | 01-JAN-1998 | (TREMblrel. 05, Last sequence update) | |
| DE | 01-MAR-2002 | (TREMblrel. 20, Last annotation update) | |
| DS | Muc1 (Fragment). | | |
| OS | Oryctolagus cuniculus (Rabbit). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus. | | |
| OX | NCBI_TaxID:9986; | | |
| RV | [1] | | |
| RP | SEQUENCE OF 410-602 FROM N.A. | | |
| RC | STRAIN=NEW ZEALAND WHITE. | | |
| RX | MEDLINE=96351712; PubMed=8747930. | | |
| RA | Spicer A.P., Dubig T., Chilton B.S., Gendler S.J., | | |
| RT | "Analysis of mammalian MUC1 genes reveals potential functionally | | |
| RT | important domains"; | | |
| RL | Mamm. Genome 6:885-888(1995). | | |
| RP | [2] | | |
| RC | SEQUENCE FROM N.A. | | |
| RC | STRAIN=NEW ZEALAND WHITE. | | |
| RA | Hewelson A., Chilton B.S.; | | |
| RL | "Molecular cloning and hormone-dependent expression of rabbit Muc1 in | | |
| RL | the cervix and uterus". | | |
| RL | Biol. Reprod. 0:0-0(1997). | | |
| RP | [3] | | |
| RC | SEQUENCE FROM N.A. | | |
| RC | STRAIN=NEW ZEALAND WHITE. | | |
| RA | Hewelson A., Chilton B.S.; | | |
| RL | Submitted (JAN-1997) to the EMBL/GenBank/DBJ | | |
| RL | databases. | | |
| RR | EMBL; 085787; AAB64380.1; | | |
| RR | Interpict IP8000082; SEA_domain. | | |
| RR | Pfam; PF01390; SEA.1. | | |

| DR | SMART; SM00200; SEA; 1. | DR | SMART; SM00204; SEA; 1. |
|---|--|---|--|
| DR | PROSITE; P550024; SEA; 1. | DR | PROSITE; P550024; SEA; 1. |
| PT | NON TER | PT | NON TER |
| SEQ | SEQUENCE | SEQ | SEQUENCE |
| 602 AA; | 61287 MW; 0F4523CF2871E270 CRC64; | 602 AA; | 61287 MW; 0F4523CF2871E270 CRC64; |
| Query Match | 48.2%; Score 1234; DB 6; Length 602; | Query Match | 48.2%; Score 1234; DB 6; Length 602; |
| Best Local Similarity | 54.7%; Pred. No. 1.7e-67; | Best Local Similarity | 54.7%; Pred. No. 1.7e-67; |
| Matches 283; Conservative 42; Mismatches 136; Indels 56; Gaps 10; | | Matches 283; Conservative 42; Mismatches 136; Indels 56; Gaps 10; | |
| OY | 22 TGSCHASSPTGGEKETSATQSSVPS-----STREKNAVMSVSV-----LSH 64 | OY | 22 TGSCHASSPTGGEKETSATQSSVPS-----STREKNAVMSVSV-----LSH 64 |
| DB | 99 TSSSTVSDDISPVHEVTSASATSPVHEVTSALTTSPVSVATSPVHEVTSAPATSPVSAT 158 | DB | 99 TSSSTVSDDISPVHEVTSASATSPVHEVTSALTTSPVSVATSPVHEVTSAPATSPVSAT 158 |
| OY | 65 SP-----GGSSPTGQODYTAAPAFPSGSAATGQVTSVPTP----- 106 | OY | 65 SP-----GGSSPTGQODYTAAPAFPSGSAATGQVTSVPTP----- 106 |
| DB | 159 SVHEVTSAPATSPVSATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVSVAT 218 | DB | 159 SVHEVTSAPATSPVSATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVSVAT 218 |
| OY | 107 -----ALGSTPPADVTSAPDNKRPAFGSTAPPAAGVTSAPDTPRPFGSTAPPAAGVTS 161 | OY | 107 -----ALGSTPPADVTSAPDNKRPAFGSTAPPAAGVTSAPDTPRPFGSTAPPAAGVTS 161 |
| DB | 219 PVHEVTSASATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVSVATSPVHEVTS 278 | DB | 219 PVHEVTSASATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVSVATSPVHEVTS 278 |
| OY | 162 PDTPRPAAGSTAPPAAGVTSAPDNKRPAFGSTAPPAAGVTSASGSA--SGSASTLVANGTSA 219 | OY | 162 PDTPRPAAGSTAPPAAGVTSAPDNKRPAFGSTAPPAAGVTSASGSA--SGSASTLVANGTSA 219 |
| DB | 279 LTTPSPVTSATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVTSATSPVHEVTS 338 | DB | 279 LTTPSPVTSATSPVHEVTSAPATSPVSVATSPVHEVTSAPATSPVTSATSPVHEVTS 338 |
| OY | 220 RATTTP--ASKTPPSPBHSDDPTTLASHSTKTDASVTHSHVTPPLTSANSTSPQLST 278 | OY | 220 RATTTP--ASKTPPSPBHSDDPTTLASHSTKTDASVTHSHVTPPLTSANSTSPQLST 278 |
| DB | 339 LATTPMPGVLSTSPSVPS--HTGPTPLTSPS-----HSTGSPVTTSTOSTPQVSA 388 | DB | 339 LATTPMPGVLSTSPSVPS--HTGPTPLTSPS-----HSTGSPVTTSTOSTPQVSA 388 |
| OY | 279 GVSEFFPSPFINLQFNSLSDPSSTDYQGLORDISEMPLQIVYQSGFGGLNIFRPGS 338 | OY | 279 GVSEFFPSPFINLQFNSLSDPSSTDYQGLORDISEMPLQIVYQSGFGGLNIFRPGS 338 |
| DB | 389 GLSFFPSPFINLQFNSLSDPSSTDYQGLORDISEMPLQIVYQSGFGGLNIFRPGS 448 | DB | 389 GLSFFPSPFINLQFNSLSDPSSTDYQGLORDISEMPLQIVYQSGFGGLNIFRPGS 448 |
| OY | 339 VVVOVLTAPEGTINVDVETQVQNKTEASRNNTIDVSHVHPFSPSAGAVPG 398 | OY | 339 VVVOVLTAPEGTINVDVETQVQNKTEASRNNTIDVSHVHPFSPSAGAVPG 398 |
| DB | 449 VVVOVLTAPEGTINVDVETQVQNKTEASRNNTIDVSHVHPFSPSAGAVPG 505 | DB | 449 VVVOVLTAPEGTINVDVETQVQNKTEASRNNTIDVSHVHPFSPSAGAVPG 505 |
| OY | 339 WQIALLVNVLVALATVYLIALAACCCRRVNGQDIPRAPDTHMSRYTHORY 458 | OY | 339 WQIALLVNVLVALATVYLIALAACCCRRVNGQDIPRAPDTHMSRYTHORY 458 |
| DB | 506 WQIALLVNVLVALATVYLIALAACCCRRVNGQDIPRAPDTHMSRYTHORY 565 | DB | 506 WQIALLVNVLVALATVYLIALAACCCRRVNGQDIPRAPDTHMSRYTHORY 565 |
| OY | 459 VPSPSTDRSPTEKVSAGNGSSLSITNPAAVATSNL 495 | OY | 459 VPSPSTDRSPTEKVSAGNGSSLSITNPAAVATSNL 495 |
| DB | 566 VPSPSTDRSPTEKVSAGNGSSLSITNPAAVATSNL 602 | DB | 566 VPSPSTDRSPTEKVSAGNGSSLSITNPAAVATSNL 602 |
| RESULT 3 | | RESULT 3 | |
| Q8MWL4 | PRELIMINARY; PRT; 580 AA. | Q8MWL4 | PRELIMINARY; PRT; 580 AA. |
| AC | O8MWL4; | AC | O8MWL4; |
| DT | 01-MAR-2002 (TEMBLrel. 20, Created) | DT | 01-MAR-2002 (TEMBLrel. 20, Created) |
| DT | 01-MAR-2002 (TEMBLrel. 20, Last sequence update) | DT | 01-MAR-2002 (TEMBLrel. 20, Last sequence update) |
| DT | 01-JUN-2002 (TEMBLrel. 21, Last annotation update) | DT | 01-JUN-2002 (TEMBLrel. 21, Last annotation update) |
| DE | MUC1 protein precursor. | DE | MUC1 protein precursor. |
| OS | Bos taurus (Bovine). | OS | Bos taurus (Bovine). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | OC | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; |
| OC | Bovidae; Bovine; Bos. | OC | Bovidae; Bovine; Bos. |
| CC | NCBI_TaxID=9913; | CC | NCBI_TaxID=9913; |
| RN | [1] | RN | [1] |
| RP | SEQUENCE FROM N. A. | RP | SEQUENCE FROM N. A. |
| RC | STRBAN-HOLSTEIN DAIRY COW; TISSUE=LACTATING MAMMARY GLAND; | RC | STRBAN-HOLSTEIN DAIRY COW; TISSUE=LACTATING MAMMARY GLAND; |
| RX | MEDLINE=21672491; PubMed=1814015; | RX | MEDLINE=21672491; PubMed=1814015; |
| RA | Pallisen L.F., Andersen M.H., Nielsen R.L., Berglund L., | RA | Pallisen L.F., Andersen M.H., Nielsen R.L., Berglund L., |
| RA | Rasmussen L.K., Petersen T.B., Rasmussen J.G.; | RA | Rasmussen L.K., Petersen T.B., Rasmussen J.G.; |
| RT | Purification of MUC1 from bovine milk-fat globules and | RT | Purification of MUC1 from bovine milk-fat globules and |
| RL | "J. Dairy Sci. 84:2591-2598 (2001)." | RL | "J. Dairy Sci. 84:2591-2598 (2001)." |
| EMBL | EMBL; AJ400824; CAC81810.1; - | EMBL | EMBL; AJ400824; CAC81810.1; - |
| OR | Interpro; IPR000082; SEA_domain | OR | Interpro; IPR000082; SEA_domain |
| OR | Pfam; PF01390; SEA; 1. | OR | Pfam; PF01390; SEA; 1. |
| OR | SMART; SMO0200; SEA; 1. | OR | SMART; SMO0200; SEA; 1. |

DR PROSITE, PS50024; SEA, 1.
 KW Signal. 1 22 MUC1 PROTEIN.
 FT SIGNAL. 23 580
 SQ SEQUENCE 580 AA; 58091 MW; E91C13984F7D757 CRC64;
 Query Match 48.0%; Score 1228.5; DB 6; Length 580;
 Best Local Similarity 48.4%; Pred. No. 3.5e-67;
 Matches 285; Conservative 44; Mismatches 157; Indels 103; Gaps 11;

QY 1 MTPGDSPPFLILL---LTVLTVTGSGHASTPGCEKEKATQKRSVSTKNAVSM 56
 DB 1 MTPDIOAPFLILLFVLITVANVPTLTSTINPRRTTPVSTQSSPTKSTWSMT 60
 QY 57 TSSVLSHSPGSGSTQGDVTLAPATEPAGSAAITWQGDVTSVPTRAL----- 108
 DB 61 TTTLLTASSPASPAPSPGHDASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPG 120
 QY 109 -GSTP-----PAHDVTSAPDKKPAKPGSTAPAHGVTSAPTRRPPGSTA--- 152
 DB 121 DGASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPG 180
 QY 153 -----PAA---HGVTSAPTRRPAPGSTAPAHGVTSAPTRRPPGSTAPVH 196
 DB 181 DGASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPG 240
 QY 197 NVTSA-----SGSAS 206
 DB 241 NGTSSPTGSPAPSPAPSPGHDASTPTSPAPSPAPSPGHDASTPTSPAPSPAPSPG 300
 QY 207 GSASTLVHNGTSARATTPPASKSTPSPSPHSDPTTLASHTKTDASSTHSTVPLT 266
 DB 301 SSWTSAKKGTSARATTPVSKGTSPSPSPHSDPTTLASHTKTDASSTHSTVPLT 355
 QY 267 SSNHSSTPOLSTGVSFPFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLF 326
 DB 356 SSNPKTSQSLSVSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLF 415
 QY 327 LGLSNKRPSPSVVLTAFREGTINHDVEOPNOYTKAASVYNTTISDVSHVAF 386
 DB 416 LGLSEIKRPSPSVVLTAFREGT-TAEWKAFQSLAHNAS-YNTTISGVSNAPF 473
 QY 387 PPSAOSGAGVGMGIALVYCVYVALAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLA 446
 DB 474 PPSAOSGAGVGMGIALVYCVYVALAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLA 533
 QY 447 SERTVTHGTVSPSTSP 495
 DB 534 SEISTVTHGTVSPSTSP 580

RESULT 4
 Q99K60 PRELIMINARY; PRT, 631 AA.
 AC Q99K60;
 DT 01-JUN-2001 (TREMblrel, 17, Created)
 DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
 DE Mucin 1, transmembrane.
 GN MUC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005441; AAH05441.1; -
 DR MGI: 97231; Muc1.
 DR InterPro: IPR000082; SEA-domain.
 DR Pfam: PF01390; SEA, 1.
 DR SMART: SM00200; SEA, 1.

DR PROSITE, PS50024; SEA, 1.
 KW Transmembrane. 631 AA; 64690 MW; B4236DE6655F4C3 CRC64;
 SQ SEQUENCE 631 AA;
 Query Match 46.8%; Score 1198; DB 11; Length 631;
 Best Local Similarity 44.8%; Pred. No. 2.8e-65;
 Matches 284; Conservative 64; Mismatches 144; Indels 142; Gaps 13;

QY 1 MTPGDSPPFLILLITVL---TVVTGSGHAST-----PG 32
 DB 1 MTPGIDAPFLILLILALIKGFLALPSEBNSVTSQDTSSILASTTPVHSSNDPATRP 60
 QY 33 GKEKTSATGRSSVSTKNAVMTSVLSGSH-----PGSGSTTQGDVTLAPATEP 86
 DB 61 GDSSTPVSOSTSPATRAPEDSTTAVLSTGSPATTA PVNASSPVANGDTSSPATSL 120
 QY 87 A--SGSAAITWQGDVTSVPTRALGSTTP-AHDVTSAPDNKPAKPGSTAPAHGVTSA 143
 DB 121 SKDSNSPVHSGTSSAPATTA PVNDSTSPVHGGTLLSPATSPGDDSTSPDHSSTSPA 180
 QY 144 TRPPGST-----APAHGVTSAPTRRPAGS-TAPPAGVTSA 181
 DB 181 TRAPEDSTTAVLSTGSPATTA PVNDSTSPVHGGTLLSPATSPGDDSTSPDHSSTSPA 240
 QY 182 PDNRPALGSTAPVH---VTSAGSA- 205
 DB 241 PATSPLRDSTSPVHSGASIONIKTSDLASTPDHNGTSVTTSSALGASATSPDHGTS 300
 QY 206 -----SGSASTLVHNGTSARAT----- 222
 DB 301 TTNSSSDLATTPVYSSMPSTTKVTSGLIIPDHNGSVLPTSSVLGASATSLVNTSAI 360
 QY 223 -TPPAKSTPSPISHHSDPTTLASHTKTDASSTHSTVPLTSSNHSSTPOLSTGV 281
 DB 361 ATPVNGTQSPSPSPVSPATTAATSSHTSIASSSVYSPFSTSSNS-SQPLVGV 419
 QY 282 FFFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLF 341
 DB 420 FFFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLFSLF 478
 QY 342 QLTARREGTIVHDVEOPNOYTKAASRYNLTISDVSHVPPFPFAGSGAGVGMGI 401
 DB 419 ESTVFRREGTIVHDVEOPNOYTKAASRYNLTISDVSHVPPFPFAGSGAGVGMGI 537
 QY 402 ALVIVCVYVALAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLA 461
 DB 538 ALVIVCVYVALAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLAIVYLA 597
 QY 462 SSTDRSPYKVSAGNGSSLSSTNPAVAATSANL 495
 DB 598 GSTRSPYKVSAGNGSSLSSTNPAVAATSANL 631

RESULT 5
 Q95L89 PRELIMINARY; PRT, 554 AA.
 AC Q95L89;
 DT 01-DEC-2001 (TREMblrel, 19, Created)
 DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
 DE Mucin (Fragment).
 GN MUC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Sacchi P.; Raesato R.; Rogati S.;
 RL Molecular analysis of MUC1 polymorphism in cattle.
 DT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF399757; AAL28023.1; -.

```
DR InterPro: IPR000082; SEA_domain.
DR Pfam: PF01390; SEA; 1.
DR PROSITE: PSS0024; SEA; 1.
FT NON TER 554
SQ SEQUENCE 554 AA; 55508 MW; E441CD140093AFD CRC64;

Query Match 43.7%; Score 1120.5; DB 6; Length 554;
Best Local Similarity 46.7%; Pred. No. 1.3e-60;
Matches 262; Conservative 41; Mismatches 157; Indels 101; Gaps 10;

QY 1 MTPGSGSPFFLLT-----LVTVVTVGSGASSTPGCKETSAIORSVPSSTKENVSM 56
DB 1 MTPDIAQAPFLSLLLPVLTVANVPLTSDSINFRRTTPVTSSPSSPTKETSMST 60
QY 57 TSSVLSHSGSGSGSTTGODVTLAPTEPASGSAATWGQVTVSVPTPAL----- 108
DB 61 TTTLLTASPPAPSPAPSPCHDGAFTTSSPAPSPAPSGHDGASTTSSPAPSPAPSGH 120
QY 109 -GSTP-----PAHDVTSAPDNKEAPGSTAPPAHGVTSAPDTPRPPGSTA---- 152
DB 121 DGASTPTSSPAPSPAPSPCHDGAFTTSSPAPSPAPSGHNGTSSPTSSPAPSPAPSGH 180
QY 153 -----PAA---HGVTAPDTPRAGSTAPPAHGVTSAPDNKPAIGSTAPPV 196
DB 181 DGASTPTSSPAPSPAPSPCHGNGTSSPTGSPAPSPAPSGHDGASTTSSPAPSPAPSGH 240
QY 197 NVTSK-----SSSAS 206
DB 241 NGTSSPTGSPAPSPAPSPCHDASLTSSPAPSPAPSPCHGASPTSSDTSMTTSSMS 300
QY 207 GSASTLVHNGTSARATTPPAKSTPPSPISRHSDPTTLASHTKPDASTHSTHVPPLT 266
DB 301 SSMVTSAHKTTSSRAATTPVSKGTPSSVPS--SRAAPMAASHITRTAPAS--PSIALST 355
QY 267 SSNNSTSPOLSTGVSFFLSPHISNLOPNSLEDSCTGYOELORISMPLOYKGGCP 326
DB 356 SSNPRTQGLSVVSLVFLSPRINLOPNSLENGCTGYOELORISMPLOYKGGCP 415
QY 327 LGLSNIKFRGSGVYVQTLAARECTINHDVETQNOVKYTAASRNLTISVYSVHP 386
DB 416 LGLSEIKFRGSGVYVETLAREGT-PAEWYKAGSQAELAHMS-YNLTLSGVYSYAPF 473
QY 387 PPSNOSAGVPGMGIALVLCVVALATVYLITAAVCCRRKNGQLDIPADTYHEM 446
DB 474 PPSNOSAGVPGMGIALVLCVVALATVYLITAAVCCRRKNGQLDIPADTYHEM 533
QY 447 SEVPTHTGRVYPPSTDRG 467
DB 534 SEVPTHTGRVYPPSTDRG 554

RESULT 6
ID 060551 PRELIMINARY; PRT; 193 AA.
AC 060551.
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Eutelestomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
```

```
DR EMBL: L41545; AAB48541.1; -.
DR InterPro: IPR000082; SEA_domain.
DR Pfam: PF01390; SEA; 1.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PSS0024; SEA; 1.
FT NON TER 1
SQ SEQUENCE 193 AA; 21201 MW; 835CF01D27BC46BA CRC64;

Query Match 28.1%; Score 720.5; DB 11; Length 193;
Best Local Similarity 70.9%; Pred. No. 8.5e-37;
Matches 139; Conservative 25; Mismatches 29; Indels 3; Gaps 3;

QY 300 DPSDYYOELORISMEFLQYKGGFLGLSNIKFRGSGVYVQTLAARECTINHDVET 359
DB 1 DSSNVYQELKRVSGLEFLQYFSR-APLGISTIFRSGSVVDVITFREKAVNASEVKS 59
QY 360 QFNQYKTEAASRYNLTISDVSVHPPEPSAOSGAGVPGMGIALVLCVVALATVYL 419
DB 60 QILCHEOE-AEYNALISKIVGEMQPPSSAOSMGVPGMGIALVLCVVALATVYL 118
QY 420 ALAVCCRRKNGQLDIPADTYHMSSEPTHTGRVYPPSTDRSPYKVSAGNGSG 479
DB 119 ALAVCCRRKNGQLDIPADTYHMSSEPTHTGRVYPPSTDRSPYKVSAGNGSG 177
QY 480 SLSTYNAVAATSANL 495
DB 178 SLSTYNAVAATSANL 193

RESULT 7
ID 028078 PRELIMINARY; PRT; 192 AA.
AC 028078.
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Eutelestomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;
RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
DR EMBL: L41543; AAB48538.1; -.
DR InterPro: IPR000082; SEA_domain.
DR Pfam: PF01390; SEA; 1.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PSS0024; SEA; 1.
FT NON TER 1
SQ SEQUENCE 192 AA; 20930 MW; 01DFFA618079715E CRC64;

Query Match 27.9%; Score 716; DB 6; Length 192;
Best Local Similarity 74.5%; Pred. No. 1.6e-36;
Matches 146; Conservative 12; Mismatches 34; Indels 4; Gaps 3;

QY 300 DPSDYYOELORISMEFLQYKGGFLGLSNIKFRGSGVYVQTLAARECTINHDVET 359
DB 1 NQTSYVQELORSINGLITQYKQDFLGLSEIKFRGSGVYVETLAREGT-PAEWYKA 59
QY 360 QFNQYKTEAASRYNLTISDVSVHPPEPSAOSGAGVPGMGIALVLCVVALATVYL 419
DB 60 QPSOLEHAAS-YNLTISGVSVYSAFPSSAQAQGVPGMGIALVLCVVALATVYL 118
QY 420 ALAVCCRRKNGQLDIPADTYHMSSEPTHTGRVYPPSTDRSPYKVSAGNGSG 479
DB 420 ALAVCCRRKNGQLDIPADTYHMSSEPTHTGRVYPPSTDRSPYKVSAGNGSG 479
```



```

Db 119 ALVCCCRKKGGLDIFPRLDHYHPMSESYTHYTHGRVVPGSTKSPYEVSAGNGS 178
QY 480 SLSTYNPAVAATSAANL 495
Db 179 NLSYTN--LAATSAANL 192

RESULT 8
Q28723 PRELIMINARY; PRT; 193 AA.
ID Q28723
AC Q28723;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Dunig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains."
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41544; AAB48540.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FR NON_TER
SQ SEQUENCE 193 AA; 21069 MW; POC31DA4D98C80B9 CRC64;

Query Match 27.6%; Score 706.5; DB 6; Length 193;
Best Local Similarity 73.3%; Pred. No. 6, 1e-36;
Matches 143; Conservative 15; Mismatches 34; Indels 3; Gaps 2;

QY 301 PESTDYQELQDISEMFLQYKQGFGLSNIKFRPGSVVQLTLAFBGTINVDHVT 360
Db 2 PSEKTYOELQNVASALISQYIGKXFLSGIRKRSSEVVLILAFBGSSTHEVTRSQ 61
QY 361 PNOYKTEASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALLVLCVLAIVYL 420
Db 62 LQNIPOAA-RINLMSRYVADVLSFSSAQS--GVPGMGIALLVLCVLAIVYL 118
QY 421 LAVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPGSTKSPYEVSAGNGSS 480
Db 119 LAVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPGSTKSPYEVSAGNGSSG 178
QY 481 LSTYNPAVAATSAANL 495
Db 179 LSTYNPAVAATSAANL 193

RESULT 9
Q60408 PRELIMINARY; PRT; 193 AA.
ID Q60408
AC Q60408;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Cavia cutleri (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognath; Caviidae; Cavia.
OX NCBI_TaxID=10144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;

```

```

RX MEDLINE=96351712; PubMed=8747930;
RA Spicer A.P., Dunig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains."
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41544; AAB48540.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FR NON_TER
SQ SEQUENCE 193 AA; 21661 MW; D5593E4EB9ECL2F CRC64;

Query Match 27.3%; Score 699.5; DB 11; Length 193;
Best Local Similarity 70.9%; Pred. No. 1, 6e-35;
Matches 139; Conservative 20; Mismatches 34; Indels 3; Gaps 3;

QY 300 PESTDYQELQDISEMFLQYKQGFGLSNIKFRPGSVVQLTLAFBGTINVDHVT 359
Db 1 PESTRYQELQENITRLFLQIYQO-DFLGLNIFKRGSAVESTVIFRNNAVESEFVS 59
QY 360 QENOYKTEASRYNLTISDVSVSHVPPPSAQSAGVPGMGIALLVLCVLAIVYL 419
Db 60 QLTQK-EDATRYMLVISEVSAEYQVT--SSTAQSGVPGMGIALLVLCVLAIVYL 117
QY 420 ALAVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPGSTKSPYEVSAGNGSS 479
Db 118 ALAVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPGSTKSPYEVSAGNGG 177
QY 480 SLSTYNPAVAATSAANL 495
Db 178 SLSTYNPAVAATSAANL 193

RESULT 10
Q35770 PRELIMINARY; PRT; 74 AA.
ID Q35770
AC Q35770;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE Mucin 1 (Fragment).
GN MUC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96285070; PubMed=9623612;
RA Descura M.M., Mahi S.K., Julian J., Carson D.D.;
RT "Reduction of mucin-1 expression during the receptive phase in the rat
RT uterus."
RL Biol. Reprod. 58:1503-1507(1998).
DR EMBL; AF00754; AAB62948.1; -.
FR NON_TER
SQ SEQUENCE 74 AA; 8187 MW; 12255F84740D9EAF CRC64;

Query Match 14.0%; Score 358; DB 11; Length 74;
Best Local Similarity 85.1%; Pred. No. 3, 4e-15;
Matches 63; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 422 AVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPGSTKSPYEVSAGNGSSL 481
Db 1 AVCCCRKKGGLDIFPARDTHPMSEYPTHTHGRVVPATTKSPYEVSAGNGSSGL 60
QY 482 SYTNPAVAATSAANL 495
Db 61 SYTNPAVAATSAANL 74

RESULT 11
Q92B39

```

```

ID Q92B39 PRELIMINARY; PRT; 487 AA.
AC Q92B39;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Multiple banded antigen.
OS Ureaplasma urealyticum (Ureaplasma urealyticum biotype 2).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
NCBI_TaxID=2130;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng X., Teng L.-J., Watson H.L., Glass J.I., Cassell G.H.;
RT "Sequence analysis of the major serovar specific antigen (M8 antigen)
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50459; AAD0075.1; -.
DR InterPro; IPR003302; Cornifin.
DR Pfam; PF02389; Cornifin; 4.
SQ SEQUENCE 487 AA; 46665 MW; B901866DADB3724 CRC64;

Query Match 10.4%; Score 267.5; DB 2; Length 487;
Best Local Similarity 32.5%; Pred. No. 9,7e-09;
Matches 100; Conservative 28; Mismatches 117; Indels 63; Gaps 16;

QY 3 PCTGSPFLLLLLVTVVSSGHSASTGCKET-----SATGRSSVSGSTERNANVMT 57
DB 149 POSGS-----TTPGSG-STTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 195
QY 58 SSVLSHSPGSGSTTGODVTLAP-----ATPAGSAGATGQDVTSP---VTPPALG 109
DB 196 -----QPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 248
QY 110 STTPPAADVTSAPDN-----XPAGSTAPPAHGVTSAP---DTPPGSTAPAAHGVTS 161
DB 249 STTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 308
QY 162 P-----DTPPAHGVTSAPDN-----RPAIGSTAPPVHNVTSASGSA-----GSA 209
DB 309 PGGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 368
QY 210 STLVHNGTSARATTPPAKSTPPSPISPHSHDPTPTLASHSTKTDASTH---HSTVPLLT 266
DB 369 ST-----TTPGSGSTTPGSGST---TTPGSGSTTPGSGSTTPGSGSTTPGSGSTTPGSG 421
QY 267 SSNHSTSP 274
DB 422 GSGSTTP 429

RESULT 12
Q9RKR9 PRELIMINARY; PRT; 1334 AA.
AC Q9RKR9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative multi-domain regulatory protein.
GN SC02359 OR SC075A.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

EN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach H., Kleser H.M., Denaplatte D., Richner A., Cullum J.,
RA Kinsch H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RP [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Lark L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz B., Rajandream M.A., Rutherford K., Puter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nucleic Acids Res. 30:141-147(2002).
DR EMBL; AL133201; CAB61705.1; -.
DR InterPro; IPR005158; BAD.
DR InterPro; IPR000767; Disease_regist.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF03704; BAD; 1.
DR PRINTS; PR00364; DISBASERSIST.
DR PROSITE; PRO1574; TUBBIFPROTEIN.
DR Prodom; PD000329; Trans_reg_C; 1.
SQ SEQUENCE 1334 AA; 138787 MW; 78DC74683E8778C CRC64;

Query Match 10.4%; Score 265.5; DB 16; Length 1334;
Best Local Similarity 33.7%; Pred. No. 4.1e-08;
Matches 91; Conservative 14; Mismatches 90; Indels 75; Gaps 12;

QY 31 PGEKETSATGRSSVSGSTERNANVMTSSVLSHSPGSGSTTGODVTLAPATEPAGS 90
DB 237 PPELRRTMAELLS-PSPTPTGSRSTPTGTPGSGAGAAAGCTDV-----ASGA 287
QY 91 AATWGDVTSVPTRPAL-----GSTTPPAHGVTSAPDNKPAAG- 129
DB 288 GAAGCPDPAAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 347
QY 130 -----STAPPAHGVTSAPDTR--PPPGS-----TAPPAHGVTSAPDTRPAGSTAP- 173
DB 348 TSPAGTAPAAAGTAPAAAGTAPAAAGTAPAAAGTAPAAAGTAPAAAGTAPAAAGTAPAAAG 406
QY 174 -----PAHGVTSAPDNRPALGSTAPPVHNVTSASGSA-----SGSASTLVH 214
DB 407 AAGTAPVAGTTPPAAGTAPAGST--PAPGTVPAPGTAPAGPAGPAGPAGPAGPAGPAGPAGS 460
QY 215 NCTSAATTPP--ASKSTPPSPISPHSDPP 242
DB 461 -GTGGAATPPPAAGAAAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 489

RESULT 13
Q9N4S7 PRELIMINARY; PRT; 1079 AA.
AC Q9N4S7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Y51B11A.1 protein.
GN Y51B11A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
NCBI_TaxID=6235;
RN [1]

```

[illegible]

Best Local Similarity 9.9%; Score 254.5; DB 6; Length 160;
 Matched 60; Conservative 14; Mismatches 82; Indels 7; Gaps 3;

```

QY      71 STTGCGVTLAATPAPASGAATWGDVTSVPTRPALGCTTPPAHVNTSADPNKAPAGS 130
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db      3 AASPDHGASTPTSSPASPSPALPGHDASTTTSSAPSPALSGDHGASTTSSPASP 62
QY      131 TAPPACGTSAEDTPPPGSTAPAAHGVTSAPDTPRAPCSTTPAPACGTSAEDNRPALGS 190
Db      63 AASGDHGAATPTSSPASPSPALPGHNTSSPTGSPAPASPAGSDHGAATPTSSPASP 122
QY      191 TAPPVHNVTSA--SGS-ASGSASTLVHNGTSARATTTPASKRNP 231
Db      123 AASFGHNTGTSPTGSPASPSPALPGHDAS----TPTSSPAP 160
RESULT 15
Q9BKRV7 PRELIMINARY; PRT; 1325 AA.
ID AC Q9BKRV7
OC Q9BKRV7
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE PT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PG3.
GN PF33.
NC Leishmania major.
CC Eukaryote; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
NCBI_TaxId=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Myler P.J., Slik E., Cawthra J., Handley F., Vogt C., Robertson L.,
   McDonald P., Ivens A., Nguyen D., Minden H., Stuart K., Worthey E.A.;
   Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC084329; AAC13175.1; -.
DR InterPro; IPRO01611; LRR.
DR InterPro; IPRO03591; LRR_typ.
DR Pfam; PF00560; LRR_8.
DR PRINTS; PRO0019; LEURICHPT.
DR SMART; SM00369; LRR_TYP; 5.
SQ SEQUENCE 1325 AA; 13385 MW; 8898928BAE9A18F CRC64;

Query Match          9.9%; Score 253.5; DB 5; Length 1325;
Best Local Similarity 27.2%; Pred. No. 2.2e-07;
Matches 97; Conservative 65; Mismatches 143; Indels 51; Gaps 12.
```

```

QY      27 ASSTPGEKETATGRSSSVSTERNAVSMYSY--LSHSPPSGSSTTGQDVTLAPAT 84
       :|::||::||::||::||::||::||::||::||::||::||::||::||::
Db      834 SSAPSASASSSAPSSSSSAPSSSSAPSSSSSAPASSSSAPASSSSAPSSSABSS 893
QY      85 SPASGSAATWGDVTSVPTRPALGCTTPPAHDVTAP---DNKPARGSTAHPAGVTS 140
       :|::||::||::||::||::||::||::||::||::||::||::||::||::
Db      894 SSSSAPASASSSAPSSSSSAPASSSSAPSSSSSAPASGSAAPSSSSSAPASSSS 953
QY      141 ADPRPPPGSTAPAAHGVTSAP---DTPAPGCTAAPAHGVTSAP---DNRPALGSTAP 193
       :|::||::||::||::||::||::||::||::||::||::||::||::||::
Db      954 APSS---SSSAPASNS--SSAPSSSSSAPASSSSAPSSSSSAPASSSSAPSSSSAP 1008
QY      194 PVHNVTASGASGASATLVHNGTSARAATTTPASKRNP-----SISSHSDPTPLAS 247
Db      1009 -----SASSSAPSSSS---SAPASSSGASAPSSSSSAPASSSSAPSSSSAPSSS 1059
QY      248 --HSKTVDASTHTSTVTPPLTSSNMHTSPQLSTGYEFFLSHISNIQPISSLPESTDY 305
Db      1060 APSSSSSAPASASSSAPSSSAPASSSAPSSSSTSLSGSGDGCHSDV----PY 1114
QY      306 YDELORDISEMPLQ-----IYKGSGFLGLSNIKFRPGSGVVQLTLAFREG 350
Db      1115 YTPPVASHNRFLSGFSSEPADASATWCCKPFCSMGILTCRPEGFSFEELGGVVKSG 1170
```

Search completed: May 1, 2003, 23:12:47

Thu May 8 16:14:05 2003

us-09-658-621b-2.rpt

Page 8

Job time : 110.089 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:30 ; Search time 17.6786 Seconds
(without alignments)
1161.337 Million cell updates/sec

Title: US-09-658-621B-2

Sequence: 1 MPTGTQSPFFLLLTTLVTV.....NGGSLSYTNPAVAANTSANL 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 2234 | 87.2 | 475 | 1 MUC1_HYLLA | Q29435 hylobates 1 |
| 2 | 2130 | 83.1 | 1255 | 1 MUC1_HUMAN | P19941 h mucin 1 p |
| 3 | 1207.5 | 46.9 | 630 | 1 MUC1_MOUSE | O02496 mus musculus |
| 4 | 1156.5 | 45.1 | 676 | 1 MUC1_MESAU | O60528 mesocricetu |
| 5 | 236.5 | 9.2 | 907 | 1 VGP3_BRV | P03300 epeirino-bar |
| 6 | 234 | 9.1 | 1162 | 1 TONA_TRYCR | P23253 trypanosoma |
| 7 | 233.5 | 9.1 | 1367 | 1 AMVH_YEAST | P08640 saccharomyc |
| 8 | 229 | 8.9 | 3178 | 1 YS89_CAEU | O09624 caenorhabdi |
| 9 | 225 | 8.8 | 886 | 1 VGP3_BRV28 | O07284 epeirino-bar |
| 10 | 221 | 8.8 | 725 | 1 AGA1_YEAST | P32323 saccharomyc |
| 11 | 216.5 | 8.5 | 797 | 1 VGLX_HYER | P28968 equine hept |
| 12 | 209.5 | 8.1 | 5179 | 1 MUC2_HUMAN | O02817 homo sapien |
| 13 | 207.5 | 8.1 | 1161 | 1 DAM4_YEAST | P47179 saccharomyc |
| 14 | 206.5 | 8.1 | 605 | 1 YHAC_YEAST | P19467 mus musculus |
| 15 | 205 | 8.1 | 605 | 1 YHAC_YEAST | P38739 saccharomyc |
| 16 | 200 | 7.8 | 610 | 1 MUC4_HUMAN | O10168 echinoscch |
| 17 | 189.5 | 7.4 | 551 | 1 HODA_BABIT | O09102 homo sapien |
| 18 | 186 | 7.3 | 1152 | 1 MAF4_HUMAN | O28645 oycyclagius |
| 19 | 183.5 | 7.2 | 1159 | 1 MAF4_HUMAN | P27816 bovine sapov |
| 20 | 182 | 7.1 | 400 | 1 MAF1_XETA | P52581 leetus apov |
| 21 | 180 | 7.0 | 634 | 1 HMF1_GALZ | P16583 xerops 11e |
| 22 | 180 | 7.0 | 634 | 1 VMLG_GALZ | P16583 xerops 11e |
| 23 | 180 | 7.0 | 634 | 1 VMLG_GALZ | P16583 xerops 11e |
| 24 | 176 | 6.9 | 721 | 1 VMS0_HSV1 | O10130 icdactid1 |
| 25 | 176 | 6.9 | 721 | 1 VMS0_HSV1 | O10130 icdactid1 |
| 26 | 176 | 6.9 | 1083 | 1 TSD3_HUMAN | O00268 homo sapien |
| 27 | 176 | 6.9 | 1229 | 1 N121_HUMAN | O09143 homo sapien |
| 28 | 175 | 6.8 | 530 | 1 YAP4_SCHRO | O09188 echinoscch |
| 29 | 174 | 6.8 | 1115 | 1 EATN_SORBI | P24352 sorghum bic |
| 30 | 174 | 6.8 | 1115 | 1 MAF1_MOUSE | P13395 mus muscula |
| 31 | 174 | 6.8 | 1115 | 1 MAF1_MOUSE | P27346 mus muscula |
| 32 | 174 | 6.8 | 1140 | 1 TM96_YEAST | O04893 saccharomyc |
| 33 | 174 | 6.8 | 1185 | 1 DRPL_HUMAN | P54259 homo sapien |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 173.5 | 6.8 | 1251 | 1 YQJ3_CAEU | O09550 caenorhabdi |
| 35 | 172.5 | 6.7 | 1072 | 1 MAF1_BOVIN | P16225 bos taurus |
| 36 | 171 | 6.7 | 636 | 1 YKX2_YEAST | P16225 bos taurus |
| 37 | 170.5 | 6.7 | 528 | 1 FOD4_HUMAN | O00592 homo sapien |
| 38 | 169.5 | 6.6 | 526 | 1 NUG2_MOUSE | O63850 mus musculu |
| 39 | 169 | 6.6 | 1150 | 1 APMD_PIG | P12021 sus scrofa |
| 40 | 168 | 6.6 | 817 | 1 VRF1_YEAST | P17270 saccharomyc |
| 41 | 168 | 6.6 | 1589 | 1 PHP_DROME | P33763 drosophila |
| 42 | 167.5 | 6.5 | 378 | 1 LKXK_RAT | P13368 ratius noii |
| 43 | 167 | 6.5 | 600 | 1 SP96_DICDI | P13328 dictyostell |
| 44 | 166 | 6.5 | 567 | 1 CH13_CANAL | P40934 canidia alb |
| 45 | 166 | 6.5 | 786 | 1 STUE_DROME | Q05319 drosophila |

ALIGNMENTS

| RESULT 1 | ID | MUC1_HYLLA | STANDARD: | PRT: | 475 AA. |
|-------------|--|--------------------------------|-------------|----------------------------|-------------|
| AC | Q29435 | 15-JUL-1999 (Rel. 38, Created) | | | |
| DT | 15-JUL-1999 (Rel. 38, Last sequence update) | | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | | |
| DE | Mucin 1 precursor (MUC-1). | | | | |
| GN | MUC1. | | | | |
| OS | Hylobates lar (Common gibbon). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates. | | | | |
| OX | NCBI_TaxID=9580; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=9635172; PubMed=8747930; | | | | |
| RA | Spicer A.P., Dubig T., Chilton B.S., Gendler S.J.; | | | | |
| RT | "Analysis of mammalian MUC1 genes reveals potential functionally | | | | |
| RT | important domains." | | | | |
| RL | Mamm. Genome 6:885-888(1995). | | | | |
| CC | - FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN | | | | |
| CC | - SUBCELLULAR LOCATION: Type 1 membrane protein. | | | | |
| CC | - PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED. | | | | |
| CC | - SIMILARITY: CONTAINS 1 SEA DOMAIN. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL: L41589; AAA69965.1; - | | | | |
| DR | EMBL: L41625; AAA69918.1; - | | | | |
| DR | EMBL: L41624; AAA69918.1; JOINED. | | | | |
| DR | InterPro: IPR000082; SEA_domain. | | | | |
| DR | Pfam: PF01390; SEA; 1. | | | | |
| DR | SMART: SM00200; SEA; 1. | | | | |
| DR | PROSITE: PS50024; SEA; 1. | | | | |
| KW | Glycoprotein; Signal; Cytokeleton; Actin-binding; Transmembrane; | | | | |
| KW | Repeat. | | | | |
| FT | SIGNAL | 1 | 23 | POTENTIAL. | |
| FT | CHAIN | 24 | 475 | MUCIN 1. | |
| FT | DOMAIN | 24 | 380 | EXTRACELLULAR (POTENTIAL). | |
| FT | TRANSMEM | 381 | 401 | POTENTIAL. | |
| FT | DOMAIN | 402 | 475 | CYTOPLASMIC (POTENTIAL). | |
| FT | REPEAT | 102 | 121 | 1. | |
| FT | REPEAT | 122 | 141 | 2. | |
| FT | REPEAT | 142 | 161 | 3. | |
| FT | REPEAT | 162 | 181 | 4. | |
| FT | DOMAIN | 254 | 371 | SEA. | |
| SQ | SEQUENCE | 475 AA; | 49371 MW; | D7A699D6D8C6622 CRC64; | |
| Query Match | | 87.2%; | Score 2234; | DB 1; | Length 475; |

Best Local Similarity 89.1%; Pred. No. 1.3e-100;
Matches 441; Conservative 8; Mismatches 26; Indels 20; Gaps 1;

QY 1 MTPGTOGPPPLLLLTTCVTVTGSGHASTPGGKEKETSATQSSVSPSTERNNAVMTSSV 60
D 1 MTPGTOGPPPLLLLTTCVTVTGSGHASTPGGKEKETSATQSSVSPSTERNNAVMTSSV 60
QY 61 LSSHPGSGSGSTTGQGVTLAPATERPAGSAAATWQGVTVVTPRPAAGSTTPPADVTS 120
D 61 LSSHPGSGSGSTTGQGVTLAPATERPAGSAAATWQGVTVVTPRPAAGSTTPPADVTS 120
QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTPAAHGVTSAPDTRPAAGSTAPPAHGVTS 180
D 121 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTPAAHGVTSAPDTRPAAGSTAPPAHGVTS 180
QY 181 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTPAAHGVTSAPDTRPAAGSTAPPAHGVTS 240
D 181 APDNKPAPGSTAPPAHGVTSAPDTRPPGSTPAAHGVTSAPDTRPAAGSTAPPAHGVTS 240
QY 241 TPTTLASHTKTDASTHSTHSTVPPPLTSSNHSSTPQSTGVSPFFELSPHISMLQFNSSLED 300
D 241 TPTTLASHTKTDASTHSTHSTVPPPLTSSNHSSTPQSTGVSPFFELSPHISMLQFNSSLED 300
QY 301 PSTDYVGELOQDISEMFLQIYKQGGFLGSLNKKFPGSVVVOULTAFREGTTNHDVETQ 360
D 301 PSTDYVGELOQDISEMFLQIYKQGGFLGSLNKKFPGSVVVOULTAFREGTTNHDVETQ 360
QY 361 PNQKTEAASRYNLITISVSVSHVPPFSAQAGVPMQIALVLCVLAVALATVYLLA 420
D 361 PNQKTEAASRYNLITISVSVSHVPPFSAQAGVPMQIALVLCVLAVALATVYLLA 420
QY 421 LAVQCCRRKRYGQLDIPARPTYHMPSEYPTHTHGRVPPSSTDRSPYKVSAGKSGSS 480
D 421 LAVQCCRRKRYGQLDIPARPTYHMPSEYPTHTHGRVPPSSTDRSPYKVSAGKSGSS 480
QY 481 LSTYNPAAVTAASNTL 495
D 481 LSTYNPAAVTAASNTL 495
QY 495 LSTYNPAAVTAASNTL 495
D 495 LSTYNPAAVTAASNTL 495

RESULT 2
MUC1_HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P15942; P13931; P17626; Q14128; Q16437; Q9Y4J2;
AC Q16615; Q14876; Q9UE75; Q9UE76; Q9UOL1; Q9BXA4;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23Ag) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RN SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Pancreas;
RA MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Bata S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA";
RL J. Biol. Chem. 265:15294-15299 (1990).
RN (2)
RN SEQUENCE FROM N.A. (ISOFORMS A AND B).
RA MEDLINE=90202794; PubMed=2318825;
RA Ligenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilken J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini";
RL J. Biol. Chem. 265:5573-5578 (1990).

RN [3]
R SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peet N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin";
RL J. Biol. Chem. 265:15286-15293 (1990).
RN [4]
RN SEQUENCE FROM N.A. (ISOFORM A).
RA MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peet N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit";
RL Biochem. Biophys. Res. Commun. 173:1019-1029 (1990).
RN [5]
RN SEQUENCE FROM N.A. (ISOFORM SRC).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms";
RL Eur. J. Biochem. 189:463-473 (1990).
RN [6]
RN SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RA MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue";
RL Eur. J. Biochem. 189:475-486 (1990).
RN [7]
RN SEQUENCE FROM N.A. (ISOFORM A).
RA MEDLINE=91031045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen";
RL Gene 93:313-318 (1990).
RN [8]
RN SEQUENCE FROM N.A. (ISOFORM Y).
RA MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Barnuch A., Eliroy-Stein O., Sagiv D.,
RA Keydar I., Hilken J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue";
RL Eur. J. Biochem. 224:787-795 (1994).
RN [9]
RN SEQUENCE FROM N.A. (ISOFORMS X; Y AND Z).
RA MEDLINE=97355747; PubMed=9212288;
RA Oosterkamp H.M., Scheiner T., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/2)";
RL Int. J. Cancer 72:87-94 (1997).
RN [10]
RN SEQUENCE FROM N.A. (ISOFORM Y).
RA Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RN SEQUENCE FROM N.A. (ISOFORM S).
RA Tissue=Epithelial cancer;
RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=88330762; PubMed=3417635;
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
 RA Burchell J.,
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 RT expressed by carcinomas is made up of tandem repeats.";
 RL J. Biol. Chem. 263:12820-12823(1988).
 RN [13]
 RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM B).
 RX MEDLINE=9008473; PubMed=2597151;
 RA Abe M., Siddiqui J., Kufe D.,
 RT "Sequence analysis of the 5' region of the human DF3 breast
 RT carcinoma-associated antigen gene.";
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 RN [14]
 RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM B).
 RX TISSUE=Thyroid;
 RA MEDLINE=96183746; PubMed=8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.,
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 RT transcriptase polymerase chain reaction of the MUC1 gene.";
 RL Int. J. Cancer 66:55-59(1996).
 RN [15]
 RP SEQUENCE OF 1-89 FROM N.A.
 RX TISSUE=Lung;
 RA MEDLINE=96181716; PubMed=8604237;
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
 RA Lee J.N., Luh K.T., Wu C.W.,
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 RT tissues.";
 RL Oncology 53:118-126(1996).
 RN [16]
 RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS C AND D).
 RX TISSUE=Breast carcinoma;
 RA Bulwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.,
 RT Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=97460054; PubMed=9312074;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 RA Hantsch F.-G.,
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 RT lactation-associated MUC1. All putative sites within the tandem
 RT repeat are glycosylation targets in vivo.";
 RL J. Biol. Chem. 272:24780-24793(1997).
 RN [18]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed=10372415;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 RA Hantsch F.-G.,
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 RT MUC1 of T47D breast cancer cells.";
 RL J. Biol. Chem. 274:18165-18172(1999).
 RN [19]
 RP POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=21359366; PubMed=11350974;
 RA Engelmann K., Belaid S.E., Hantsch F.-G.,
 RT "Identification and topology of variant sequences within individual
 RT repeat domains of the human epithelial tumor mucin MUC1.";
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [20]
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed=10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
 RA Staeler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.,
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 RT its cognate binding protein.";
 RL Cancer Res. 59:1552-1561(1999).
 RN [21]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.

RX MEDLINE=21240104; PubMed=11341784;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.,
 RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [22]
 RP CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed=11847293;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
 RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
 RA Stacey M., Lin H.-H., Gordon S.,
 RT "Generation of ligand-receptor alliances by 'SPA' module-mediated
 RT cleavage of membrane-associated mucin proteins.";
 RL Protein Sci. 11:698-706(2002).
 RN [23]
 RP PHOSPHORYLATION.
 RX MEDLINE=95080414; PubMed=7988707;
 RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.,
 RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
 RT Cytokine receptor-like molecules.";
 RL FEBS Lett. 356:130-136(1994).
 CC -1- FUNCTION: May play a role in adhesive functions and in cell-cell
 CC interactions, metastasis and signaling. May provide a protective

Query Match 83.1%; Score 2130; DB 1; Length 1255;
 Best Local Similarity 39.0%; Pred. No. 3.3e-95;
 Matches 489; Conservative 1; Mismatches 5; Indels 760; Gaps 1;

QY 1 MTPGQSPFFLLTLTLVTVGSGHASTPGEKETSA TORSSVPSSTKNAVSMTSSV 60
 DB 1 MTPGQSPFFLLTLTLVTVGSGHASTPGEKETSA TORSSVPSSTKNAVSMTSSV 60
 QY 61 LSHSPGSGSTTQGDVTLAPATEPPASGSAATMGQDVTSVVPFPALGSTTPADVTS 120
 DB 61 LSHSPGSGSTTQGDVTLAPATEPPASGSAATMGQDVTSVVPFPALGSTTPADVTS 120
 QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 164
 DB 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 164
 QY 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 180
 DB 121 APDNKPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT----- 180
 QY 165 ----- 164
 DB 165 ----- 240
 QY 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
 DB 181 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 240
 QY 165 ----- 164
 DB 165 ----- 164
 QY 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
 DB 241 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 300
 QY 165 ----- 164
 DB 165 ----- 360
 QY 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
 DB 301 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 360
 QY 165 ----- 164
 DB 165 ----- 420
 QY 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
 DB 361 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 420
 QY 165 ----- 164
 DB 165 ----- 540
 QY 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
 DB 481 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 540
 QY 165 ----- 164
 DB 165 ----- 600
 QY 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
 DB 541 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 600
 QY 165 ----- 164
 DB 165 ----- 660
 QY 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
 DB 601 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 660
 QY 165 ----- 164

```
Db 661 APTTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 720
Qy 165 ----- 164
Db 721 APTTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 780
Qy 165 ----- 164
Db 781 APTTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 840
Qy 165 ----- 164
Db 841 APTTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 900
Qy 165 ----- 200
Db 901 APTTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 960
Qy 201 AGSASAGSASTLVHNGTSARATTPASKSTPPSIPSHSDPTTLASHTKTDASTHHS 260
Db 961 AGSASAGSASTLVHNGTSARATTPASKSTPPSIPSHSDPTTLASHTKTDASTHHS 1020
Qy 261 TVPPLTSSNHSTSPOLSTGVSPFELSFHISNLQFNSSLEDPSTDIYQELORDISEMFLQI 320
Db 1021 SVPLTSSNHSTSPOLSTGVSPFELSFHISNLQFNSSLEDPSTDIYQELORDISEMFLQI 1080
Qy 321 YKGGGFLGSLNFKPPGSAVVVQTLAFREGTINVDVETQFNQYKTEAASRYNLITSPVS 380
Db 1081 YKGGGFLGSLNFKPPGSAVVVQTLAFREGTINVDVETQFNQYKTEAASRYNLITSPVS 1140
Qy 381 VSHVPPFSAAGAGVPMGIALNVCVVALAVYILALANCCCRKNYQCDIPAR 440
Db 1141 VSDPPFSAAGAGVPMGIALNVCVVALAVYILALANCCCRKNYQCDIPAR 1200
Qy 441 DTHPMSEVPTHTHGRVYPPSSTDRSPYEKVSAGNGSSLSSTNPAVANSANI 495
Db 1201 DTHPMSEVPTHTHGRVYPPSSTDRSPYEKVSAGNGSSLSSTNPAVANSANI 1255

RESULT 3
MUC1 MOUSE STANDARD; PRT; 630 AA.
ID MUC1 MOUSE AC 002436;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Episialin).
GN MUC1 OR MUC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RX MEDLINE=1714452;
RX Spicer A.P., Parry G., Paton S., Gendler S.D.;
RA "Molecular cloning and analysis of the mouse homologue of the tumor-
RT associated mucin, MUC1, reveals conservation of potential O-
RT glycosylation sites, transmembrane, and cytoplasmic domains and a
RT loss of misfolded-like polymorphism."
RT J. Biol. Chem. 266:15099-15109(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92068178; PubMed=1958179;
RA Vos H.L., Devries Y., Hilkens J.;
RT "The mouse episialin (Muc) gene and its promoter: rapid evolution of
RT the repetitive domain in the protein."
RT Biochem Biophys Res Commun. 181:121-130(1991).
CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED
CC IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY
```



```
QY 1 MPTGQSPFFLLTLTLV-----TYTGGHAST-----PG 32
DB 1 MPTGIRAFPLLLLLSLKGLALPSENAVSTSSQOTSSSLASTTTPVHSSNDPATRP 60
QY 33 GKEKTSATQSSVPSSTERNNAVSTSSVLSNH-----PGSGSTTGQDVLATPATP 86
DB 61 GUSTSSPVQSTSSPATRAPEDSTTAVLSGTSSPATPAVVASASSPVAGDSTSP 120
QY 87 A--SGSAATWGQDVTSPVTRPALGSTTPPAHDVTSAPDNKPAGSTAPPAHGVTSAPDT 144
DB 121 LKDSNASSPVHSGSSAATTAVDSTSSPVHGGTSSPATSPGDSSTSPHSGTSSPAT 180
QY 145 RPPPGST-----APAAHGVTSAP 162
DB 181 RAPEDSTTAVLSGTSSPATPAVDSTSSPVAHDDTSSPATSLSDSASSPVHAGTSSP 240
QY 163 DRRPAPGSTAPPAH-----VTSAPDNK-----PALGSTAPVHANTVS-- 200
DB 241 ATSPAPDSTSSPVHSSASIQNIKTSDLASTPBNNGTSTVTTSSALSGASPHSGTSTT 300
QY 201 -----ASGSA-----SGSASTVHNGTSPR 220
DB 301 TMSGSVLTATTPVYSSMPSTTKTSSGSAIIPDHNGSVLPTSSVLSGATSLVYN--TSAP 359
QY 221 ATTPPASKSTPPSPSHSDPTTLASHSTKTDASTHSTVPLTSSNHSSTPQLSTCV 280
DB 360 A--TTVNSNGTOPSPSQVPVSPMAITSSHSTLASSSYSTVPFSTPSSNS--SPQLSVGV 417
QY 281 SFFPLSFHISNLTQFNSLEDPTDYQELQORDISEMFLOIYKQGGFLGSLNIXKRPQSVV 340
DB 418 SFFPLFEPYIQHPNSSLSDPSSNYYQELKKNISGLFLOIF--NGDFLGISSIKFRSGSVV 476
QY 341 VOLTLAPFEGTINVDVETOFNQYTEAPSRYNLTISDVSGHVPFPPSAOSGAGVPGWG 400
DB 477 VESTVTFHEGTFSSADSVKSLQHKHEAD--YNLTISEVKNVEMQFPSSAGRGVPGWG 535
QY 401 ILALVLCVVALAIYVLAIALAVCCCRKXNGOLDTPAPADYTHMSXPTHTHRKVP 460
DB 536 IALVLCVLAIALAIYFLALAVCCCRKXSGQDLDFPQDTHHMSYPTHTHRKVP 595
QY 461 PSTDRSPYEKVASGNGSSSLYTPAVATSNAL 495
DB 596 PSTKRPQYEVSAAGSSSSLYTPAVVTSANL 630
RESULT 4
MUC1 MESAU STANDARD; PRF: 676 AA.
AC 060528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor.
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NCBI_TaxID=10036;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=tracheal epithelium;
RA MEDLINE=9632618; PubMed=8703480;
RA Park H., Hyun S.W., Kim K.C.;
RA "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
RT cells in primary culture."
RL Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DB EMBL: U36918; AAB53965.1;
DB InterPro: IPR000082; SEA_domain.
DB Pfam: PF01390; SEA; 1.
DB SMART: SM00200; SEA; 1.
DB PROSITE: PS50024; SEA; 1.
KW Glycoprotein; Signal; Cytochrome; Actin-binding; Transmembrane;
KW Repeat.
FT CHAIN 1 25
FT SIGNAL 26 676
FT DOMAIN 26 582
FT TRANSMEM 583 603
FT DOMAIN 604 676
FT CHAIN 458 573
FT CARBOHYD 291 323
FT CARBOHYD 323 350
FT CARBOHYD 350 380
FT CARBOHYD 380 400
FT CARBOHYD 400 413
FT CARBOHYD 413 435
FT CARBOHYD 435 479
FT CARBOHYD 479 496
FT CARBOHYD 496 536
SQ SEQUENCE 676 AA; 67616 MW; 95F479B6EC5C384A CRC64;
Query Match 45.1%; Score 1156.5; DB 1; Length 676;
Best local similarity 40.6%; Pred. No. 6.6e-49;
Matches 276; Conservative 61; Mismatches 154; Indels 189; Gaps 12;
QY 1 MPTGQSPFFLLTLTLV-----TYTGGHAST-----PG 32
DB 1 MPTGIRAFPLLLLLSLKGLALPSENAVSTSSQOTSSSLASTTTPVHSSNDPATRP 60
QY 25 GHASS-----TPGKEKTSATQSSVPSST-----EKNAVSTSSVLSG-- 63
DB 61 GHSGSAPPTSSAVNSATTPGHSGSSAPPTSSAVNSATTPVHSGSSAPVTSVAVDSATTP 120
QY 64 -HSPGS-----GSSTTGQDVTLPATEPAGSAAATWGQDVTSPVTRPALGSTTTP 114
DB 121 VHSGSAPPTSSAVNSATTPVHSGSSAPVTSVAVNSATTPVHSGSSAPVTSVAVDSATTP 180
QY 115 AHDVTSAPDNKPAPGSTAPPAHGVTSAPDTPPPGSTAPPAHGVTSAPDPAPGSTAPP 174
DB 181 VHSGSAPPTSSAVNSATTPVHSGSSAPVTSVAVNSATTPVHSGSSAPVTSVAVDSATTP 240
QY 175 AHGVTSAPDNKPALGSTAPVHANTVS----- 201
DB 241 VHSGSAPPTSSAVNSATTPVHSGSSAPPTSSAVNSATTPVHSGSSAPVTSVAVDSATTP 300
QY 202 -----SGSAS----- 206
DB 301 VPPGSMQTTAISGSANP IINGSLVPTTSSALVPTTSAHSGASAMTSSSDLAATP 360
QY 207 -----GSASTLVHNGTSAPATTPPAKSTPPSP 235
DB 361 IDSCTSIETKPAATTPVHNGSLVPTTSSVLSGATYLLINDSTVTAATTPVGNQTSVP 420
QY 236 SHSDPTTLASHSTKTDASTHSTVPLTSSNHSSTPQLSTGVSPFFLSFHLSNQFN 295
DB 421 SRHPPTPPPAVSSNSTIALSTYSTALSPAFSSHA-AQVAVGVSPFLSFHIMHQN 479
QY 296 SSLEDPSTDYQELQORDISEMFLOIYKQGGFLGSLNIXKRPSPVAVQTLAIFRGSTINVA 355
DB 480 SSLEDPSSNYYQELKXNVSGTLQVFSR-AVLGIETIFRSGSVVVDSTVIFRGAVNAS 538
```

QY 356 DVEIENQYKTEAASRYNLTISDVSVHPPFSSAGAGVPGMGIALLVICVLVALAI 415
 DB 539 EVKSOIHOEABEAYNIAISKINVEQFSSAQSPVPGMGIALLVICVLVALAI 597
 QY 416 VYIATAVQCCKRRKYGOLDIFPARDTYHPMSEYPTTHGRVYPPSSTRSPYKVSAG 475
 DB 598 VYIATAVQCCKRRKYGOLDIFPDOSYHPMSEYPTTHGRVYPPSSTRSPYKVSAG 657
 QY 476 NGSSSLSTYNPAVATSANL 495
 DB 658 N-GSSSLSTYNPAVATSANL 676
 RESULT 5
 VGP3_EBV STANDARD; PRT; 907 AA.
 AC P03200; P03201; (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [contains:
 GN Glycoprotein GP220].
 OS BILFI.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Segun C.,
 RA Tuffnell P.S., Barrell B.G.,
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RT Nature 310:207-211(1988).
 CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CD2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01555; CAA24854.1; --
 DR PIR: A03762; O0BR21.
 DR PIR: A03763; O0BR22.
 DR PIR: S33008; S33008.
 KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
 FT CARBOHYD 87 47 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 533 533 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 610 610 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 656 656 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 815 815 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 888 888 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPIC 502 698 MISSING (IN GP220).
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;
 Query Match 9.2%; Score 236.5; DB 1; Length 907;
 Best Local Similarity 23.1%; Pred. No. 9e-05;
 Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;
 QY 22 TGSGHAST-----PGGEKTSATORSSVPSTENAVASMTSSVLSHSPGSGSTP 73
 DB 444 TTGGLPSSTHVPNTLTPAAGTPTVST--ADVTSPTAGTSGASPTVPSPSPMDNGTES 501
 QY 74 QGQDVT--LAAPTEPASGSAATMGQDVTSV-----VTRPALGSTPPPH 116
 DB 502 KADMTSSSPVPTPTPNATSPPTAVTPTPNATSPPTAVTPTPNATSPFLGKSTPS 561
 QY 117 DVTSADP-NKAPGSTAPPAHGVTSAPD-TREPPGSTAPPAHGVTSADP-TREPPGSTAP 173
 DB 562 VTPPTNATSPFLGKSTPSAVTTPPNATSPFLGKSTPSAVTTPPNATSPFLGKSTPS 621
 QY 174 PAHG-----VTSAPDNPALGSTAPPVNAVTSAGSA-----SGSASTL-- 212
 DB 622 QANATHTLGGSTPTPVVTSQPKMATSVTGO--NHITSSSTSSMSLRPSNPTETSPS 679
 QY 213 -VHNGTSARATTPA-----SKSTPPSPSHSDPTTLASTHTDASTHST 261
 DB 680 TSDNSTSHMPLTFAHPTGGENITVTPASISTHVSSTSPAPRGTSQASGSGNSTS 739
 QY 262 -----VPELTSSNHSTSPOLSTGVSEFFPSFHSINLOPNSL-----EDP 301
 DB 740 TKPEVNVTKTPPQMATSPQAPSGQKTAVPTVSTGCKANSTGSKHTTGAGRTSTP 799
 QY 302 STDYQELRODISMFLQIKQGGFLGSLNKKPRGVSVVQTLIAFRSGTINVDVETP 361
 DB 800 TTDY-----GDSSTPRRYNATYILPST----- 824
 QY 362 NQYKTEAASRYNLTISDVSVH--VPPFSSAGAGVPGM-GIALLV-CVVALAIY 417
 DB 825 ---SSKLRPRMTFTSPPTTAQATVPVPTSQ---FRPSNLSLVQMASLAVLTLL 876
 QY 418 LIALAVQCCKRRKYGOLDIFPARDTYHPMSEYPTTHGRVYPPSSTRSPY 469
 DB 877 LLMVADCAFR-----NLSTSHT---YTPPYDAETV 906
 RESULT 6
 TCNA_TRYCR STANDARD; PRT; 1162 AA.
 ID TCNA_TRYCR
 AC P23253;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sialidase (EC 3.2.1.16) (Neuraminidase) (NA) (major surface antigen).
 GN TCNA.
 OS Trypanosoma cruzi.
 OS Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

NCBI_Taxid=5693;
 [1] SEQUENCE FROM N.A.
 RC STRAIN-S13V10 X-10/4;
 RC MEDLINE=91277609; PubMed=1711561;
 RA Pereira M.E.A., Mejia U.S., Ortega-Barria E., Matilevich D.,
 Piroli R.F.,
 RT the trypanosoma cruzi neuraminidase contains sequences similar to
 RT bacterial neuraminidases, tryd repeats of the low density lipoprotein
 RT receptor, and type III modules of fibronectin.
 RL J. Exp. Med. 174:179-191(1991).
 RM [2]
 RN SUBCELLULAR LOCATION.
 RP MEDLINE=91376547; PubMed=1896773;
 RA Piroli R.F., Mejia U.S., Ajl T., Alkawa M., Pereira M.E.A.;
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
 RT trypanomastigotes."
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
 CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
 CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
 CC PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M61732; AAA30255.1; -
 DR PIR: JH0557; JH0557.
 DR InterPro: IPR002860; GH_BNR.
 DR Pfam: PF02012; BNR_2
 KW Hydrolyase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
 KW Phosphorylation.
 FT DOMAIN 1 457
 FT REPEAT 23 34
 FT REPEAT 163 174
 FT REPEAT 209 220
 FT DOMAIN 458 588
 FT REPEAT 569 1120
 FT DOMAIN 1120 1200
 FT CARBOHYD 342 342
 FT CARBOHYD 394 394
 FT CARBOHYD 1125 1125
 FT SEQUENCE 1162 AA; 120032 MW; 0704922189706A40 CRC64;
 SQ
 Query Match 9.1%; Score 234; DB 1; Length 1162;
 Best Local Similarity 28.7%; Pred. No. 0.00015;
 Matches 87; Conservative 34; Mismatches 134; Indels 48; Gaps 11;
 QY 16 TULVTGSGH-----ASTPGKEKTSATQSSVPSSTKRNAMTSVLSHS 65
 DB 707 TPSTPVDSSAHGTPSTPADSSAHSTSTPADSSAHSTSTPADSSAHGTPSTPADSSAHG 766
 QY 66 PGSGSSITGQDVTIAPTEPASGSAT---WQDVTSVVTPRALGSTTP---AHD 117
 DB 767 TPSTPVDSSAHSTSTPADSSAHGTPSTPADSSAHSTSTPADSSAHGTPSTPADSSAHG 826
 QY 118 VTSAPNKPAPSTAP---AHGVTAP-DTRPPSTAPA---AHGVTAPDTPAPG 169

DB 827 TPSTPVDSSAHGTPSTPADSSAHSTSTPADSSAHSTSTPADSSAHGTPSTPADSSAHG 886
 QY 170 STAP-----AHGVTAPDNRPALGSTAPPYANTYSAGSGSASTLV 213
 DB 887 TPSTPVDSSAHSTSTPADSSAHSTSTPADSSAHGTPSTPD---SNAHGTPSTPADSSA 944
 QY 214 HNGSARATTTTASGTPSTPSTPHNS--DTPITLASHST-KTDSASTHST-VPLITSN 269
 DB 945 HSTSTPVDSSAHSTSTPADSSAHSTSTPADSSAHSTSTPADSSAHSTSTPADSSA 1004
 QY 270 HST 272
 DB 1005 HST 1007
 RESULT 7
 ID AMYH_YEAST STANDARD; PRT; 1367 AA.
 AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
 DE glucosidase) (1,4-alpha-D-glucan glucosylhydrolase).
 GN STAI OR STAI2 OR MAL5 OR Y1R019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentsel S., Hamlyn N., Horenell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Randeram M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RA MEDLINE=87194600; PubMed=3106330;
 RA Yamashita T., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RT STAI".
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RA MEDLINE=89031230; PubMed=3141213;
 RA Pardo J.M., Jerez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STAI2 and SGA genes
 RT from Saccharomyces cerevisiae".
 RL FEBS Lett. 239:179-184(1988).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: TO S.POMBE SPC215.13.
 CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: Z38061; CAAB6176.1; -
 DR EMBL: M16164; AAA35014.1; -
 DR EMBL: M16165; AAA35015.1; -
 DR EMBL: X13857; CAA32069.1; -
 DR PIR: B26877; B26877.

```

DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S000458; MUC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CAROXYD 817 817
FT CAROXYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 9.1%; Score 233.5; DB 1; Length 1367;
Best Local Similarity 28.9%; Pred. No. 0.0019;
Matches 88; Conservative 45; Mismatches 123; Indels 49; Gaps 12;

QY 20 VVTGSGHASTPGKEKETSATORSSVP---STPEKNVAVMTSSVLSHS---PGSGSTT 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 VTSSTSSSSAPVPTSSSTSSAPVTSSTSSAPVTSSTSSAPVPTSSSAPVPTSSSTT 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 QGQDVTIAPTEPASGSAATWG-QDVTSPVTPPALGSTPEPADVTSAPEKPPA--- 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 EESSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVPTSSSAPVPTSS 463
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 ---GSAAPAHGVT---SAPDTPPGSTAPAAAGVTSADTR---PAGSTAPPAH 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 STTESAPVTSSTSSAPVPTSSSTSSAPVTSSTSSAPVPTSSSTSSAPVPTSSSTES 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 GVTSAPEKPPALGST---APVPAHVTSAGSA---SGSASTLVHNGTSARATTPAKS- 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 523 ---SSAPVPTSSSTSSAPVTSSTSSAPVPTSSSTSSAPVPTSSSTSSAPVPTSSSAP 580
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 ---TPSPISPHSTPTTLASHT-----KTDASTHHSVTPPLTSNHS- 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 581 VPTSSSTSSAPVPTSSSTSSAPVPTSSSTSSAPVPTSSSTSSAPVPTSSSTSSAPVPT 640
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 274 POLST 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 641 PESTS 645

RESULT 8
ID YS89_CABEL STANDARD; PRT; 3178 AA.
AC 009624; Q09625; Q969D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson-Spyrat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-eb.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; 248544; CAB70192.1; -.
DR EMBL; 248582; CAB70192.1; JOINED.
DR EMBL; 248582; CAB70201.1; -.
DR EMBL; 248544; CAB70201.1; JOINED.
DR Wormbed; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat channel Tpl.
DR InterPro; IPR001024; Lipoygenase LH2.
DR InterPro; IPR000636; M-channel nlg.
DR InterPro; IPR000203; PKD Cys-rich.
DR Pfam; PF00520; Ion trans; 1.
DR Pfam; PF01477; PLAT; 1.
DR Pfam; PF01825; GPS; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 266 1196
FT DOMAIN 1105 1241
FT DOMAIN 2071 2120
FT TRANSMEM 13 30
FT TRANSMEM 51 73
FT TRANSMEM 2139 2161
FT TRANSMEM 2348 2367
FT TRANSMEM 2390 2412
FT TRANSMEM 2451 2468
FT TRANSMEM 2483 2505
FT TRANSMEM 2567 2589
FT TRANSMEM 2636 2661
FT TRANSMEM 2836 2861
FT TRANSMEM 2976 2998
FT TRANSMEM 3038 3060
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Query Match 9.1%; Score 233; DB 1; Length 3178;
Best Local Similarity 26.8%; Pred. No. 0.0047;
Matches 115; Conservative 54; Mismatches 172; Indels 88; Gaps 19;

QY 19 VVTGSGHASTPGGE---KETSATORSSVPSTEKNVAVMTSSVLSHS---PGSGS- 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 302 VTTTAMSTSTSTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 STTGQDVTIAPTEPASGSAATWGQDVTSPVTPPALGSTPEPADVTSAPEKPPA 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 PTTTPELTSTLSLSDNALCSYLDFTTSTFTTTLTST-----TEEP----- 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 TAPPAHVTSAPDTPPEPGSTAPAAAGVTSADTRPAPGSTAPPAAGVTSAPDNKPPAGS 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 409 TETTTTETVTSST---VTTEPTTLTSTTAST---STTEPSTSTVTSP-----S 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 TAPPAHVTSAGSAGSASTLVHNGTSARAT--TPAS---KSTPST---PSHSDT 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 455 TSPVTSVTSSTSS---SSSTTVTTPPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 242 PTTTASHSTYTDASTHST---VPPLTSNHSSTPOLSTGVGFPL-----SFHIS- 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 510 PSTSTSSVSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 NLOFNSP-----EDPSDYVOELORDISEMFLQ-----IYQGGFGLS 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 570 NTLNLSGLIGIGTISIECTSPSTSNVSTTKD-CACPTKSVAMPRLGCTYASTPVGPG 628
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 NIKPPPSGV---VOLTLA---PREGTINVDVETQFNQYKTEAASRYNLISDVSVSH 383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 629 NTFPAATMTDDKVVYTVANVYIQEVSSTTESSTSSAVASTSTSTSTSTSTSTSTSTST 688
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 384 VPPPSAQS 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 689 VTEPSTRS 697

RESULT 9
ID VGP3_BEVAB STANDARD; PRT; 866 AA.

```

AC Q07284; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (NA).
 GN BLUF1.
 OS Epstein-Barr virus (strain AG876) (Human herpesvirus 4), and
 OS Epstein-Barr virus (strain P3HR-1) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 NC NCB1_TaxID=62830, 82829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AG876, and P3HR-1;
 RC MEDLINE=93311716; PubMed=8393237;
 RA Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,
 RA Arrand J.R.;
 RA "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
 RT conserved between virus types A and B.";
 RL Virology 195:578-586(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P3HR-1;
 RL Klein K., Mueller-Lantsch N.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L07922; AAA02783.1; -;
 DR EMBL; L07923; AAA02787.1; -;
 DR EMBL; X67776; CAA47986.1; -;
 KW Membrane; Glycoprotein; Antigen; Late protein.
 FT CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 680 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 886 AA; 92388 MW; 4394F6130DECCABA CRC64;

Query Match 8.9%; Score 229; DB 1; Length 886;
 Best Local Similarity 24.2%; Pred. No. 0.0002;
 Matches 136; Conservative 56; Mismatches 205; Indels 164; Gaps 26;

QY 28 SSTPGGKEXETATQRRS-----VPSSTKKNVAMTSSVLSHSPGS-G 69
 DB 370 SGTPSGCGNISGAPASRPFDTITVGGCTAGTAKTILITRTINATYTHKVFSAQPSSTT 429
 QY 70 SSTTGGQDQVTLAPATE---PAS-----GSAATWGDVTSVPVTRPALSGTTPPAHDVT 119
 DB 430 TSPITNTTGFAPAPPTTGTGPPSSHTVPTNLTPAPSTGPTVSTADVTSPPACTTSGASPVT 489
 QY 120 SAP---DN---KPAFGSIAPPAHGVTSAPD-----TRPPGSTAPAAHGVTSAPD-- 163
 DB 490 PPSPRDNGTESKAPDMWTSPTSAVTTTPFNATSPTPAVTTPFNATSPTL-GKTSPTSAV 548
 QY 164 TRPAGSTAPPAHGVTSAPD-NRPALSGTAP-----PVHNTVSAS-GSAGSGSTLVH- 214
 DB 549 TTPTRNATSPPAVTTTPFNATITPLGKTSPTSAVTTTPFNATSPVGETSPQANTNHT 608
 QY 215 -NGTSAR-ATTTPASKSTPSPSPSHSDTPTTLASHSTK-----TDASSTHST 261
 DB 609 LGGTSTPTVTSPPKATSAVTTGQHNITSSSTSMELRPSISBTLSPSTSDNSTSHM- 667
 QY 262 VPLTSSNHSSTPQSLSTGVSPFELSFHISNDQFNSLSDPSLDYQELQORDISEMPLQIY 321
 DB 668 --PLTSAHPGTGENITQVTPASTGTH--HVSTSSPAPRGTT-----TLAFRGITN- 706
 QY 322 KQCGFLGSLNKKPPGSSVYVQL-----TLAFRGITN----- 353
 DB 707 SQASGFGMSSTSTPGEVNVTYKGTTPPKATSPQAPSGOKTAVTVTSTGKANTTGAKH 766
 QY 354 -----VHDETQFNQYTEAARVRLTT---SPVSYSHVPPPF---SAQSGAG 395
 DB 767 TTGSGARSTGPTTIDYGDSTTPRRRYNATYLLPSTSSKLAPRMFTSPPTTAQATVP 826
 QY 396 VPGW-----GIALLVV-CVVALAIVYLIALAVCCRRKNYGQDIFPARDYHPMS 448
 DB 827 VPTSGPRSMVSMVLVQMASLAVLTLILLVMADCAFR-----N 867
 QY 449 YPTVTHRGRVYPPSSTDRSPY 469
 DB 868 LSTSHT---YTPPYDAETV 885

RESULT 10
 AGAL YEAST STANDARD; PRT; 725 AA.
 ID AGAL YEAST STANDARD; PRT; 725 AA.
 AC P32323;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE A-aggutinin attachment subunit precursor.
 GN AGAL OR YNR044W OR N1431.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NC NCB1_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=9104412; PubMed=2072914;
 RA Roy A., Lu C.F., Maruyama D.L., Hyde P.N., Kurjan J.;
 RT "The AGAL product is involved in cell surface attachment of the
 RT Saccharomyces cerevisiae cell adhesion glycoprotein a-aggutinin.";
 RL Mol. Cell. Biol. 11:4196-4206(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
 CC SUBUNIT. S-CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
 CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
 CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
 CC AGGREGATION DURING MATING.
 CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
 CC A CORE SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M60590; AAA34382.1; --
 CC EMBL; Z71659; CAA96325.1; --
 CC PIR; S17031; S17031.
 CC PIR; A41258; A41258.
 CC SGD; S0005327; AGA1.
 CC Glycoprotein; Cell adhesion; signal; GPI-anchor; Repeat;
 CC Pheromone response.
 CC -----
 CC FT CHAIN 1 22 POTENTIAL.
 CC FT SIGNAL 1 22 A-AGGLUTININ ATTACHMENT SUBUNIT.
 CC FT DOMAIN 23 725 2 X APPROXIMATE REPEATS.
 CC FT REPEAT 53 493 1-1.
 CC FT REPEAT 395 493 1-2.
 CC FT DOMAIN 182 307 18 X APPROXIMATE TANDEM REPEATS, SER/THR-
 CC FT REPEAT 182 188 2-1.
 CC FT REPEAT 189 195 2-2.
 CC FT REPEAT 196 202 2-3.
 CC FT REPEAT 203 209 2-4.
 CC FT REPEAT 210 216 2-5.
 CC FT REPEAT 217 223 2-6.
 CC FT REPEAT 224 230 2-7.
 CC FT REPEAT 231 237 2-8.
 CC FT REPEAT 238 244 2-9.
 CC FT REPEAT 245 251 2-10.
 CC FT REPEAT 252 258 2-11.
 CC FT REPEAT 259 265 2-12.
 CC FT REPEAT 266 272 2-13.
 CC FT REPEAT 273 279 2-14.
 CC FT REPEAT 280 286 2-15.
 CC FT REPEAT 287 293 2-16.
 CC FT REPEAT 294 300 2-17.
 CC FT REPEAT 301 307 2-18.
 CC SEQUENCE 725 AA; 73353 MW; 70420C851B0B01F8 CRC64;
 CC -----
 CC Query Match 8.8%; Score 225; DB 1; Length 725;
 CC Best Local Similarity 22.8%; Pred. No. 0.00026;
 CC Matches 94; Conservative 80; Mismatches 152; Indels 86; Gaps 15;
 CC -----
 CC QY 16 TVLVTVGSGHASSSTPGGKETSATGRSSVPSSTKNAVSMTSSVLSHSPGSGSTTQG 75
 CC DB 164 SIISPTV-SLISSTSSNPPTTSLSTSTSPST-STSPSTSTSSSTSTSSSTSS 221
 CC QY 76 QDVTLAPTEPASGAATWQDVTSVPTRPALSGSTPPAHVTSAPDKKPAFGSTAPPA 135
 CC DB 222 SSTSPSTSTSSSLTS-----TSSSSTSTSSSTSS-----TSSSSTSTSS 270
 CC QY 136 HGVTAPPTPR-PPPGSTAPPAHGVTSAP--DTPRAGSTAPPAHGVTSAP-DNKRPAIGS 190
 CC DB 271 SSTSPSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSS 330
 CC QY 191 T-----APVNAVTSAGS-----ASGASSTLVHNGTSARATTPASKSTP 231
 CC DB 331 SIASSSTSVLSYSDSTVSVPTSSNVAIPSMSTSVETTVSSQSSSYITKSIPTTI 390
 CC QY 232 PSIP-----SHSDPTTLASHST-KTDA-S-THHSTVPLTSS 268
 CC DB 391 PPSMSSTFTTVSGVTMTTMYTTCFYSSESTSTLMSHETVTTDAVCHSCMPSQTS 450
 CC QY 269 NSTSPOLSTGVSFSLFHSINLQFNLSLEDPNDYQGLQDISEMLOLYQGGFLG 328
 CC DB 451 LITSSIKST-----KVAIVSTSTVSSSTACSTCAETSHSY----- 468

QY 329 LSNIXPSPGVVOLTLAFREGTINVDVETQNOYKTEASRYNLISDVS 380
 DB 469 -SSVQVASSSVTQOTTSTKWSWMTSDDEPNKH---ATGKHVHTSSGTS 536
 DB -----
 DB RESULT 11
 DB VGLX_HSYEB STANDARD; PRT; 797 AA.
 DB AC VGLX_HSYEB
 DB P28968;
 DB 01-DEC-1992 (Rel. 24, Created)
 DB 01-DEC-1992 (Rel. 24, Last sequence update)
 DB 01-DEC-1992 (Rel. 24, Last annotation update)
 DB Glycoprotein X precursor.
 DB 71.
 DB GN Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
 DB OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 DB OC Alphaherpesvirinae; Varicelloviruses.
 DB NCBI_TaxID=31520;
 DB OK [1]
 DB RN SEQUENCE FROM N.A.
 DB RP MEDLINE=9295566; PubMed=1318606;
 DB RX Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
 DB RA "The DNA sequence of equine herpesvirus-1."
 DB RL Virology 189:304-316 (1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M8664; AAE02506.1; --
 CC PIR; H36802; VGBEX1.
 CC KW Glycoprotein; Transmembrane; Signal.
 CC FT SIGNAL 1 22 POTENTIAL.
 CC FT CHAIN 23 797 GLYCOPROTEIN X.
 CC FT DOMAIN 23 465 SER/THR-RICH.
 CC FT TRANSMEM 766 790 POTENTIAL.
 CC FT CARBOHYD 590 590 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SEQUENCE 797 AA; 80342 MW; 50C9ED9211FE5B2 CRC64;
 CC -----
 CC Query Match 8.6%; Score 221; DB 1; Length 797;
 CC Best Local Similarity 26.3%; Pred. No. 0.00044;
 CC Matches 103; Conservative 41; Mismatches 149; Indels 96; Gaps 14;
 CC -----
 CC QY 11 LLLLVTLVTVTGS-----GHASSTPGKETSAT-----TQSSVPSST 49
 CC DB 9 LLLCMVSVIYAIIGSTTTTETTTSSSSSTSGSGSTSSGTTNSSSAPPTSSPPTSS 68
 CC QY 50 EKNAVMTS-----SVLSHSPGSGSTTQ-----GQVTLAP- 82
 CC DB 69 HTSPSTSTGSSSTRAATSSSAPSTASSTTSPSTSTSTTTTTPRATTTTTPTAAPT 128
 CC QY 83 --ATEPASGAATWQDVTSVPTRPAL-GSTPRAHVTSAPDKKPAFGST-----A 132
 CC DB 129 TAATTTAVTAASTAAATTTATATATPTTTTPTTTTATTTTPTTASTTTTDTTAA 188
 CC QY 133 PPAHGVTSAPDTPRPGSTAPPAHGVTSAPDTPRPGSTAPPAHGVTSAPD----- 183
 CC DB 189 TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 248
 CC QY 184 ---NRPAGSTAPRVHNVTSAGSAGSAS-----TVHNGTSARATTT--TPASKS 229
 CC DB 249 TTAATTTAATTTGSPSTGSGSTTGTGASTSPASATATATPTSTGTAATTTPTPTSA 308
 CC QY 230 T-----PSPITPHSDPR-----TLASHSTKTDASSTHSGVPLTSSNHSSTPS 277
 CC DB 309 TSAESTTEAPSTPTPTDTPPSAATTAATTSSTVSSASTTSATTAFTTSH-TSPDSS 367
 CC QY 278 TGVSFPLFHSINLQFNLSLEDPSTDY 306

| | | |
|-----------|--|---------------|
| Dh | 368 TGST--STAEPSTFTLTPSTATSPDOF | 394 |
| RESULT 12 | | |
| ID | MUC2 HUMAN | PRT; 5179 AA. |
| AC | Q02817; Q14678; | |
| DT | 01-JUN-1994 (Rel. 29, Created) | |
| PT | 01-NOV-1997 (Rel. 35, Last sequence update) | |
| DE | Mucin 2 precursor (Intestinal mucin 2). | |
| GN | MUC2 OR SMUC. | |
| OS | Homo sapiens (human). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| OX | NCBI_TaxID=9606; | |
| RP | SEQUENCE FROM N.A. | |
| RX | TISSUE=Intestine; | |
| RA | Gum J.R., Jr., Hicks J.W., Tortibara N.W., Siddiki B., Kim Y.S., J. Biol. Chem. 267:21375-21383(1992). | |
| RT | "The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region." | |
| RL | J. Biol. Chem. 267:21375-21383(1992). | |
| RN | [3] | |
| RP | SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A. | |
| RX | MEDLINE=913016075; PubMed=1400449; | |
| RA | Gum J.R., Jr., Hicks J.W., Tortibara N.W., Rothe E.-W., Lagace R.E., Kim Y.S., J. Biol. Chem. 267:21375-21383(1992). | |
| RT | "The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region." | |
| RL | J. Biol. Chem. 267:21375-21383(1992). | |
| RN | [3] | |
| RP | SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A. | |
| RX | MEDLINE=91358717; PubMed=1885763; | |
| RA | Tortibara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W., Rothen G.M., Kim Y.S., and polyomorphism. Intestinal mucin gene structure. Repeated arrays of polymorphic repeat units. | |
| RT | "FUNCTION: COMES THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS AGENTS AT MUCOSAL SURFACES." | |
| CC | -1- SUBUNIT: MULTIMERIC. | |
| CC | -1- SUBCELLULAR LOCATION: Secreted. | |
| CC | -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS, | |
| CC | Bronchus, Cervix and Gall bladder. | |
| CC | -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR | |
| CC | INTERCHAIN DISULFIDE BONDS (BY SIMILARITY). | |
| CC | -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND | |
| CC | VARIABLE AMONG DIFFERENT ALLELES. | |
| CC | -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT | |
| CC | OF SILKWORM HEMOCYTIN. | |
| CC | -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE (CTCK) DOMAIN. | |
| CC | -1- This SWISS-Prot entry is copyright. It is produced through a collaboration | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Collaboration - | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | |
| CC | use by non-profit institutions as long as its content is in no way | |
| CC | modified and this statement is not removed. Usage by and for commercial | |
| CC | entities requires a license agreement (see http://www.isb-sib.ch/announce/ | |
| CC | or send an email to license@isb-sib.ch). | |
| DR | EMBL, L21998; AAB95295.1; | |
| DR | EMBL, W74027; AAB59875.1; | |
| DR | EMBL, M94131; AAA59163.1; | |
| DR | EMBL, M94132; AAA59164.1; | |
| DR | Genebank, HGNC:7512; MUC2. | |
| DR | MTM, I5B370; | |
| DR | InterPro: IPRO00359; Cys_knot. | |
| DR | InterPro: IPRO00561; BGF-like. | |
| DR | InterPro: IPRO02400; GF_cytnk. | |
| DR | InterPro: IPRO02919; TIF_Cystrich. | |
| DR | InterPro: IPRO01007; VWF_C. | |
| DR | InterPro: IPRO01846; VWF_D. | |
| DR | Pfam: PF00093; Cys_knot; 1. | |
| DR | Pfam: PF00097; Cys_knot; 1. | |
| DR | Pfam: PF00094; vwd; 4. | |
| DR | Pfam: PF00094; vwd; 4. | |
| DR | Pfam: PF01826; TIF; 1. | |
| DR | PRINTS: PR00438; GFCYSKNOT. | |
| DR | SMART: SM00214; VMC; 2. | |
| DR | SMART: SM00216; VMD; 4. | |
| DR | PROSITE: PS00022; BGF_1; UNKNOWN_1. | |
| DR | PROSITE: PS01185; CTCK_1; 1. | |
| DR | PROSITE: PS01225; CTCK_2; 1. | |
| DR | PROSITE: PS01208; VWF_C; 2. | |
| KW | Glycoprotein; Repeat; Signal. | |
| FT | SIGNAL | 1 20 |
| FT | CHAIN | 21 5179 |
| FT | DOMAIN | 1401 1747 |
| FT | REPEAT | 1401 1416 |
| FT | REPEAT | 1417 1432 |
| FT | REPEAT | 1433 1448 |
| FT | REPEAT | 1449 1464 |
| FT | REPEAT | 1465 1471 |
| FT | REPEAT | 1472 1478 |
| FT | REPEAT | 1479 1494 |
| FT | REPEAT | 1495 1517 |
| FT | REPEAT | 1518 1533 |
| FT | REPEAT | 1534 1556 |
| FT | REPEAT | 1557 1572 |
| FT | REPEAT | 1573 1596 |
| FT | REPEAT | 1597 1612 |
| FT | REPEAT | 1613 1631 |
| FT | REPEAT | 1632 1651 |
| FT | REPEAT | 1652 1675 |
| FT | REPEAT | 1676 1683 |
| FT | REPEAT | 1684 1699 |
| FT | REPEAT | 1700 1715 |
| FT | REPEAT | 1716 1731 |
| FT | REPEAT | 1732 1747 |
| FT | DOMAIN | 4815 4886 |
| FT | DOMAIN | 4924 4991 |
| FT | DOMAIN | 5075 5160 |
| FT | DISULFID | 5089 5132 |
| FT | DISULFID | 5098 5132 |
| FT | DISULFID | 5152 5160 |
| FT | DISULFID | 5102 5164 |
| FT | DISULFID | 5135 5169 |
| FT | CARBOHYD | 1 53 |
| FT | CARBOHYD | 425 620 |
| FT | CARBOHYD | 770 770 |
| FT | CARBOHYD | 894 894 |
| FT | CARBOHYD | 894 894 |
| FT | CARBOHYD | 1139 1139 |
| FT | CARBOHYD | 1154 1154 |
| FT | CARBOHYD | 1215 1215 |
| FT | CARBOHYD | 1230 1230 |
| FT | CARBOHYD | 1246 1246 |
| FT | CARBOHYD | 1287 1287 |
| FT | CARBOHYD | 1820 1820 |
| FT | CARBOHYD | 4339 4339 |
| FT | CARBOHYD | 4351 4351 |
| FT | CARBOHYD | 4362 4362 |

```

FT CARBOHYD 4881 4881 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1351 1351 H > L (IN REF. 3).
FT CONFLICT 1412 1412 T > S (IN REF. 3).
FT CONFLICT 1449 1449 L > P (IN REF. 3).
FT CONFLICT 1504 1504 M > T (IN REF. 3).
FT CONFLICT 4192 4192 G > S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD1571F89A5663 CRC64;

Query Match
Best Local Similarity 26.0%; Score 216.5; DB 1; Length 5179;
Matches 82; Conservative 46; Mismatches 126; Indels 61; Gaps 13;

OY 2 TGTGSPFLILLLVTVTGSGHASTGGEKETSATGSSV-----PESTKNAVMT 57
Db 1478 SEPTTTP-----SEPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1330
OY 58 SEVLSSHSRSGSS--TTCGQDVTIAPTEBPASGAATWGDVTSVPTPALGTT-- 112
Db 1531 TTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1586
OY 113 -PPADVTASAP---DNKPAPGSTAPPAHGVTSAPDTP--PQGSTA-----PAA 155
Db 1587 SEPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1646
OY 156 HGVTSAPDTPRPGST--APPAGVTSAP---DNKPALGSTAPPAHNVTSASGSASGS 208
Db 1647 TTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1706
OY 209 ASVLVHNGTSKATTPPASKSPPSIPSHSDPTPLASHTKDS--STHSVTPPLT 266
Db 1707 -----SPTTTTTSSSTTP-----SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1751
OY 267 SSMHSTSPQLSTGV 281
Db 1752 TSSPLTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1766

RESULT 13
DAN4_YEAST STANDARD; PRT; 1161 AA.
ID DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
GN DAN4 OR Y0R1S1C OR J2223.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 11
RA SEQUENCE FROM N.A.
RA Pearce T.;
RN Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RP REGULATION.
RX MEDLINE=2111166; PubMed=11160904;
RA Cohen B.D., Serflin O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites."
RL Nucleic Acids Res. 29:799-808(2001).
CC -1- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1 / T1P1 FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49651; CA99684.1;
CC SCD; S0003812; DAN4;
CC Item: P0000992; SRP1_T1P1.
CC P0000992; SRP1_T1P1.
CC PROSITE: P800724; SRP1_T1P1.
CC Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
CC STGNAL
CC CHAIN 1
CC PROPEP 1147 1146 CELL WALL PROTEIN DAN4.
CC LEIPI 1146 1146 GPI-ANCHOR (POTENTIAL).
CC SEQUENCE 1161 AA; 118358 MW; 7954C15069F0CA58 CRC64;

Query Match
Best Local Similarity 25.6%; Score 208.5; DB 1; Length 1161;
Matches 81; Conservative 46; Mismatches 148; Indels 41; Gaps 8;

OY 16 TVLVVTSGHASTPGGEKETSATGSSVPESTKNAVMTSSVLSSHSRSGSSTTG 75
Db 123 TSTSTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 178
OY 76 QDVTIAPTEBPASGAATWGDVTSVPTPALG--STTPPAHGVTSAPDNKPAPGSTAP 133
Db 179 TPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 236
OY 134 PAHGVTSAPDTPRPGSTAPPAHGVTSAPDTPRPGSTAPPAHGVTSAPDNKPAPGSTAP 193
Db 237 STTSQTSKSTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 291
OY 194 PPAHNVTSASGSASSTLVHNGTSARATTPPASKSPPSIPSHSDTP-----TT 244
Db 292 TSTSTSTSTSTSSASASVI--STTAATSTTPASLTTPATSTADTHITSSVSTNAFTT 349
OY 245 LASHSTKTD-----ASTHSTVPLT-----SMHSTSPQLSTGVSPFLSP 287
Db 350 SATTTTSDIYSSSPSQVTSAPPTVSEVTSVSEPTSSQVTSAPPTVSEPTSSV 409
OY 288 HISNLQFNLSLEDPST 303
Db 410 EPTRSSQVTSAPPTT 425

RESULT 14
C114_MOUSE STANDARD; PRT; 573 AA.
ID C114_MOUSE STANDARD; PRT; 573 AA.
AC P19467;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface antigen 114/A10 precursor.
GN LY64.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RA SEQUENCE FROM N.A.
RA MEDLINE=89197960; PubMed=2784793;
RX Dougherty G.J., Kay R.J., Humphries R.K.;
RT "Molecular cloning of 114/A10, a cell surface antigen containing
RT highly conserved repeated elements, which is expressed by murine
RT hemopoietic progenitor cells and interleukin-3-dependent cell
RT lines."
RL J. Biol. Chem. 264:6509-6514(1989).
CC -1- FUNCTION: MAY HAVE A POSITIVE REGULATORY ROLE IN THE CELLULAR

```



```

CC      RESPONSE TO IL-3.
CC      -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC      -1- PTM: THE SER/THR-RICH TANDDEM REPEATS MAY SERVE AS SITES OF
CC      EXTENSIVE GLYCOSYLATION.
CC      -1- PTM: THE MOTIF SER-GLY CONSERVED IN ALL 8 SER/THR-RICH REPEATS MAY
CC      SERVE AS THE SITE OF ATTACHMENT OF GLYCOSAMINOGLICAN SIDE CHAINS.
CC      -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J04634; AAA37239.1; -.
DR      PIR; A33533; A33533.
DR      MGD; MGI:103190; Lr64.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR000082; SEA_domain.
DR      Pfam; PF01390; SEA_1.
DR      SMART; SM00181; EGF_1.
DR      SMART; SM00001; EGF_Like_1.
DR      SMART; SM00200; SEA_1.
DR      PROSITE; PS00022; EGF_1; FALSE_NEG.
DR      PROSITE; PS01186; EGF_2; 2.
DR      PROSITE; PS50024; SEA; 1.
DR      GlycoProtex; Signal; Antigen; EGF-like domain; Repeat; Transmembrane.
FT      CHAIN * 1 17
FT      DOMAIN 18 573
FT      TRANSMEM 481 508
FT      DOMAIN 509 573
FT      DOMAIN 17 230
FT      REPEAT 17 43
FT      REPEAT 44 70
FT      REPEAT 71 97
FT      REPEAT 98 124
FT      REPEAT 125 151
FT      REPEAT 152 178
FT      REPEAT 179 205
FT      REPEAT 206 230
FT      DOMAIN 233 273
FT      DOMAIN 274 384
FT      DOMAIN 385 425
FT      DOMAIN 425 467
FT      DISULFID 237 248
FT      DISULFID 242 257
FT      DISULFID 259 272
FT      DISULFID 389 402
FT      DISULFID 394 408
FT      DISULFID 410 424
FT      DISULFID 429 441
FT      DISULFID 433 451
FT      CARBOHYD 453 466
FT      CARBOHYD 266 286
FT      CARBOHYD 316 326
FT      CARBOHYD 397 397
SQ      SEQUENCE 573 AA; 58701 MW; 1154C4P04E2D58A9 CRC64;

Query Match 8.1%; Score 207; DB 1; Length 573;
Best Local Similarity 25.6%; Pred. No. 0.0015;
Matches 101; Conservative 66; Mismatches 165; Indels 62; Gaps 18;

```

```

QY      118 VTSAPDNKPAKSTAPPAHGVTSAPDT--RPPGTAAPAHGVTSAPDRPAPGTAAPA 175
DB      119 VQSGSPGSSGASSTTTSSGASPTTVQSGSPSSGAS-----TTSSGASAPPT 171
QY      176 HGVTAPDNKPAKSTAPPAHGVTSAPDT--RPPGTAAPAHGVTSAPDRPAPGTAAPA 233
DB      172 TVQSGP-----SSS-QVSTTSSGASPTTVQSGSPSSGASPPGTPGASST 224
QY      234 IPHSHDPTPTLAS-HSTKTDASSTH-HSTVPLTSSNHSBPSJSTGVFFPLSHIT- 290
DB      225 VPGSGTSPDLCPNPKGTASCVKLHKAFCLEGGYNSLSSCVKGTTPPGDISM 284
QY      291 NLFNSLSDPSTDYQELORDISEWFLQIKQGGFLGLSNIKFRPGSVVQTLT----- 345
DB      285 SVSETPANLSDNSVQDELNSVTDFFETFNKTDY-----GQTVIIVKSTAPSR 335
QY      346 ---APREGT-----INVADVETOPNOYKTEEA 369
DB      336 ARSAMDATKDVSVSVNIGADTKETKSVSA 369

RESULT: 15
YHCB YEAST STANDARD; PRT; 605 AA.
AC P38739;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 63.8 kDa protein in GUT1-RIM1 intergenic region
DE precursor.
GN YHL028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8286C / AB972;
RX MEDLINE=94378003; Pubmed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faveil J., Fulton L., Gattung S., Geisel C., Kisten J.,
RA Kucada T., Hillier L., Jier M., Johnston L., Langston T.,
RA Latreille P., Louis E.J., Macri C., Wardi E., Meneses S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasis B., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaubin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR      EMBL; U11583; AB865040.1; -.
DR      PIR; S48940; S48940.
DR      SDD; S0001020; MSC4.
DR      InterPro; IPR002889; MSC.
DR      Pfam; PF01822; MSC_1.
DR      SMART; SM00321; MSC_1.
KW Hypothetical protein; Transmembrane; signal.
FT SIGNAL 1 26
FT CHAIN 27 605
FT DOMAIN 116 317
FT TRANSMEM 415 435
SQ      SEQUENCE 605 AA; 63807 MW; C9DF8CBA9553811 CRC64;

Query Match 8.1%; Score 206.5; DB 1; Length 605;

```

Best Local Similarity 24.8%; Pred.No.0.0016;
Matches 112; Conservative 66; Mismatches 179; Indels 95; Gaps 18;

```
QY 33 GKKETATQSSVPTSEKNAVSWTSSVLSHSPGSGSTTQGGDVTLAPATEPAGGAA 92
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 111 GQTLSSVQ--SVETSTESSVYSSSITSS-----SSTAIYDTTISPTLSTIS---- 158
QY 93 TWGQDVSVVTPRPAAGSTPRPADVTSA-PDNKAPGATAPAHGVTSAPDRPPGST 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 -----TPELI--TASTSTPSTDILSALPTTSTKLSTSIPT-STTSSTSTTSS 208
QY 152 APAHGVTSAPDRPPAPGATAPAHGVTSAPDNKAPAGSTAPPVHNVTSAGSAGSAST 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 TSTVSVTSSSTSTSTSTSTSTSTSTSTSTSTSSSSSTPTTTSAP-----ISTSTSTSTST 263
QY 212 LVHNGTARATTTTPASKSTPPSPSPHSDPTTTLASHSTKTDASTHSTVPELTSSNHS 271
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 STSPSTSSAPL--SSSNTPTPTSTFTTTSPTAPSPSTVYVTSSTASPTSTTTSVNIQ 321
QY 272 TSPOLS--TGVSFFPLSPHISNLQFNSLSDPSTDYCELORDISEMFLQIYKQ-GFIG 328
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 322 TSLKYSVITVTSVHTWDTVISEI-----TSRYLTMKKVIT---QVSGTLGATP 367
QY 329 LSNIKFRPGSVVQVTLAPREGTINVHDEVTQFNOYKTEAASRYNLITSDVSYSHPEPF 388
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 TSAVATTSASVGRITNNNSNTNSNTPNKSTKKGWDSPGKIATPVV----- 420
QY 389 SAQSGAGVPGMGIALIVYCVLVALAIVYLIALAVQCQRK----NYGQDLIFARDTY 443
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 -----GVCCLVITCIL-----IYLIHRYKTRPARKQDFENEYQSKFY----- 458
QY 444 HPMSERP-----TYHTHGRVVPDSSTDRSPY 469
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 --QSKYPNEVYTTTLHT-----PSPSNSSTF 482
```

Search completed: May 1, 2003, 23:10:54
Job time : 24.6786 secs

Thu May 8 16:14:04 2003

ue-09-658-621b-2.rpr

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:08:21; Search time 52.0536 Seconds

(without alignments)
914.184 Million cell updates/sec

Title: US-09-658-621b-2

Sequence: 1 MPTGTQSPFFLLLTTLTV.....NGSSLSLTNPVAATSANL 495

Scoring table: BLOSUM62

Gapop 10.0, Gapept 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

PIR 73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|--------|--------------------|
| 1 | 2087 | 81.5 | 1344 | 1 | A35175 | mucin 1 precursor, |
| 2 | 1899 | 62.4 | 1347 | 2 | S10571 | mucin 1 precursor, |
| 3 | 1203.5 | 46.9 | 630 | 2 | A39344 | tumor-associated m |
| 4 | 1201 | 46.8 | 631 | 2 | I32257 | episialin - mouse |
| 5 | 1174 | 45.8 | 355 | 1 | S58146 | mucin 1 precursor, |
| 6 | 564.5 | 22.0 | 356 | 2 | A60533 | tumor-associated a |
| 7 | 265.5 | 10.4 | 1334 | 2 | I50568 | probable multi-dom |
| 8 | 240 | 9.4 | 377 | 2 | A48018 | mucin 7 precursor, |
| 9 | 237 | 9.3 | 1032 | 2 | I34433 | hypothetical prote |
| 10 | 237 | 9.3 | 2232 | 2 | I34433 | hypothetical prote |
| 11 | 236.5 | 9.2 | 907 | 1 | Q08221 | membrane antigen g |
| 12 | 234 | 9.1 | 1162 | 2 | U10557 | exo-alpha-stilidas |
| 13 | 233.5 | 9.1 | 1367 | 1 | S48478 | glucan 1,4-alpha-g |
| 14 | 233 | 9.1 | 796 | 2 | T21460 | hypothetical prote |
| 15 | 233 | 9.1 | 3570 | 2 | I45025 | mucin MUC5B, trach |
| 16 | 232.5 | 9.1 | 786 | 2 | T16509 | hypothetical prote |
| 17 | 231.5 | 9.0 | 528 | 2 | I47141 | gastric mucin (C10 |
| 18 | 229 | 8.9 | 886 | 2 | S29605 | glycoprotein 350/2 |
| 19 | 225.5 | 8.8 | 886 | 2 | T46707 | proteohosphoglyca |
| 20 | 225.5 | 8.8 | 825 | 2 | T29634 | hypothetical prote |
| 21 | 225 | 8.8 | 729 | 2 | S33363 | mucin 5AC (clone J |
| 22 | 225 | 8.8 | 725 | 2 | A41258 | a-agglutinin core |
| 23 | 222 | 8.7 | 534 | 2 | T39903 | serine-rich protei |
| 24 | 221 | 8.6 | 797 | 1 | VG8E1 | glycoprotein X pre |
| 25 | 220.5 | 8.6 | 1952 | 2 | T48814 | hypothetical prote |
| 26 | 219.5 | 8.6 | 1275 | 2 | T33369 | hypothetical prote |
| 27 | 219 | 8.5 | 317 | 2 | S55316 | mucin (clone PCM-2 |
| 28 | 218.5 | 8.5 | 322 | 2 | A53715 | apomucin precursor |
| 29 | 216.5 | 8.5 | 3020 | 2 | A43932 | mucin 2 precursor, |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 215.5 | 8.4 | 4776 | 2 | E95206 | cell wall surface |
| 31 | 213 | 8.3 | 867 | 2 | T45463 | membrane glycoprot |
| 32 | 212.5 | 8.3 | 866 | 2 | T45462 | membrane glycoprot |
| 33 | 210 | 8.2 | 605 | 2 | T33913 | hypothetical prote |
| 34 | 210 | 8.2 | 2187 | 2 | T33926 | naeent polypeptid |
| 35 | 209 | 8.2 | 292 | 2 | S24169 | mucin 5A169 |
| 36 | 209 | 8.2 | 660 | 2 | U10067 | mucin 5A169 |
| 37 | 208.5 | 8.1 | 1161 | 2 | S57180 | chitinase IEC 3.2. |
| 38 | 207 | 8.1 | 573 | 2 | A33523 | probable membrane |
| 39 | 206.5 | 8.1 | 507 | 2 | A33523 | cell surface glyco |
| 40 | 206.5 | 8.1 | 605 | 2 | S48940 | antifreeze glycopo |
| 41 | 205 | 8.0 | 598 | 2 | T38403 | hypothetical prote |
| 42 | 204.5 | 8.0 | 43 | 2 | S35724 | probable nucleopor |
| 43 | 201 | 7.8 | 354 | 2 | T46740 | MUC1 protein - hum |
| 44 | 200 | 7.8 | 610 | 2 | PM0012 | microfilament shea |
| 45 | 199.5 | 7.8 | 2761 | 2 | T21064 | mucin 4, tracheal |
| | | | | | | hypothetical prote |

ALIGNMENTS

RESULT 1
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin; nreatic mucin; polymorphic epithelial mucin (PEM)
N:Contents: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor, C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000
C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S5 R. Liljtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A>Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene en
A:Reference number: A35175; MUID:90202794; PMID:2318825
A:Accession: A35175
A:Molecule type: mRNA
A:Residues: 1-952, 1033-1344 <LIG1>
A:Cross-references: GB:W32739; GB:J05288; NID:G182121; PIDN:AAA5804.1; PID:G182124; GB
A:Experimental source: splice form A
A>Note: GenBank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino-and carboxyl-e
A:Accession: B35175
A:Molecule type: mRNA
A:Residues: 1-19, 29-952, 1033-1344 <LIG2>
A:Cross-references: GB:W32739; GB:J05288; NID:G182126; PIDN:AAA5806.1; PID:G182129; GB
A:Experimental source: splice form B
A>Note: GenBank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino-and carboxyl-
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Pest, N.; Burche
J. Biol. Chem. 265, 15286-15293, 1990
A>Title: Molecular cloning and expression of a human tumor-associated polymorphic epithel
A:Reference number: A35886; MUID:90368715; PMID:1697589
A:Accession: A35886
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 29-952, 1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:G188669; PIDN:AAA5976.1; PID:G188670
A>Note: GenBank entry HUMCMB includes one copy of the tandemly repeated sequence
R:Plan, M.S.; Bacter, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A>Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716; PMID:2394722
A:Accession: A35887
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 29-1109, 5', 1111-1339, 'A', 1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:G189599
A>Note: GenBank entry HUMPMNMU contains four fewer copies of the tandemly repeated seq
R:Wreschner, D.H.; Hareven, M.; Tsataly, I.; Smorodinsky, N.; Horev, U.; Zaretzky,
Eur. J. Biochem. 189, 463-473, 1990
A>Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gene
A:Reference number: S10571; MUID:90276413; PMID:2351132
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WR>

A:Cross-references: EMBL:X52229; NID:g37053
R:Weascher, D.H.
Submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <MR>
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CA346476.1; PID:g37054
R:Abbe, M.; Siddiqui, U.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473; PMID:2597151
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142, 'O', 144-162, 'O', 164-168 <AB>
A:Cross-references: EMBL:M31823; PID:g181542; PIDN:AA35757.1; PID:g181543
R:Manuzyan, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, U.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglutinin
A:Reference number: J00235; MUID:93123189; PMID:1478915
A:Accession: J00065
A:Molecule type: mRNA
A:Residues: 996-1011, 'BS', 1014-1017, 1018-1033, 'T', 1034-1037, 1038-1057 <MA>
A:Experimental source: gastric carcinoma cell
R:Zilian-Licht, S.; Baruch, A.; Elroy-Stein, O.; Kedar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A:Reference number: S51026; MUID:95080414; PMID:7988707
A:Contents: annotation
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-q23
A:Insertions: 20/1; 62/3; 1165/3; 1164/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
F:1.1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F:1.62/Region: mucin 1 amino-terminal non-repetitive
F:1.23/Domain: signal sequence #link PDB #status predicted <SIGA>
F:1.19, 29-23/Domain: signal sequence #link PDB #status predicted <SIGB>
F:1.19, 29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F:1.19, 29-212, 1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice
F:118-1344/Region: 20-residue repeats (GSTAPAPAGVSDADPPAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:145-1372/Domain: transmembrane #status predicted <TM>
F:1046, 1064, 1118, 1144, 1222/Binding site: carboxylate (Asp) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

| Query Match | 81.5% | Score 2087 | DB 1 | Length 1344 |
|-----------------------|--|-------------------|---------------|-------------|
| Best Local Similarity | 88.0% | Pred. No. 4.4e-99 | | |
| Matches 412 | Conservative | 7 | Mismatches 47 | Indels 2 |
| | | | Gaps 1 | |
| Qy 28 | SSSTRGGEKENSANQRSSVSTENKAAVSTSSVSSSPSGSGSTTGQODVTAAPTEBA | 87 | | |
| Db 879 | STAPPAAGVTSADPTRAPAGSTAPAPAGVTSAPTRAPAG--VTSAPDTRRA | 936 | | |
| Qy 88 | SSSAATWQGVTVPTTPBALGSTTPADVYNSADNKKAPSGSTAPPAVGTSSAPDTRPP | 147 | | |
| Db 937 | PGSTAPPAAGVTSAPTRAPAGSTAPPAAGVTSADTRAPAGSTAPPAAGVTSAPDTRRA | 996 | | |
| Qy 148 | PGSTAPPAAGVTSAPTRAPAGSTAPPAAGVTSAPDNKPAFGSTAPPVHANTSSAGSAG | 207 | | |
| Db 997 | PGSTAPPAAGVTSAPTRAPAGSTAPPAAGVTSAPDNKPAFGSTAPPVHANTSSAGSAG | 1056 | | |
| Qy 208 | SASTLVANGTSARATTPTPASKSTPSPISPHSDVPTTLASHSTKTDASTHSTVPLTS | 267 | | |
| Db 1057 | SASTLVANGTSARATTPTPASKSTPSPISPHSDVPTTLASHSTKTDASTHSTVPLTS | 1116 | | |

QY 266 SNHSTFQSLSTGVSPFFSLSPHINLQFNSSLEDDSTGYVOELQORDISMFQIYKQGGCL 327
Db 1117 SNHSTFQSLSTGVSPFFSLSPHINLQFNSSLEDDSTGYVOELQORDISMFQIYKQGGCL 1176
QY 328 GLSNIKRPRPSVWVQLTAAFEQTIWHDVEVFQFNQKTEAASRNLTISVSGSVHPPE 387
Db 1177 GLSNIKRPRPSVWVQLTAAFEQTIWHDVEVFQFNQKTEAASRNLTISVSGSVHPPE 1236
QY 388 FSAQSGAGVGVGNGILALVLCVVALATYVILALAVCOCKRKNYQGLDIFPARATYHPMS 447
Db 1237 FSAQSGAGVGVGNGILALVLCVVALATYVILALAVCOCKRKNYQGLDIFPARATYHPMS 1296
QY 448 EYPTVTHGGRVPPSTSDRSPKRYKVSAGSGSSLSYTNPNVAATAMN 495
Db 1297 EYPTVTHGGRVPPSTSDRSPKRYKVSAGSGSSLSYTNPNVAATAMN 1344
RESULT 2
S10571
musn 1 precursor, secreted epithelial tumor antigen splice form - human
N:containing: Homo sapiens (man)
C:species: Homo sapiens (man)
C:date: 07-Apr-1994 #sequence: revision 07-Apr-1994 #text_change 01-Dec-2000
C:accession: S10571, JN0100, T56024, S09706, S10212
R:researcher: D.H.; Hareuveni, M.; Tearefacy, I.; Smorodinsky, N.; Horev, J.; Zaretzky, J.
Bur, J.; Biochem. 189, 463-473, 1990
A:title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gener
A:reference number: S10571, MUID:90276413; PMID:2351132
A:molecule type: mRNA
A:residues: 1-347 <WR>
A:cross-references: EMBL:X5228; NID:936434; PIDD:CA36477.1; PJD:936435
R:refatory: J.; Hareuveni, M.; Horev, J.; Zaretzky, J.; Weise, M.; Jaltetech, J.M.; Garm
Gene 93, 313-318, 1990
A:title: Isolation and characterization of an expressed hypervariable gene coding for a
A:reference number: JN0100, MUID:91033045; PMID:1688329
A:accession: JN0100
A:status: preliminary
A:molecule type: DNA
A:residues: 1-138 'CN', 137-142, 'E', 144-163, 204-208, 'A', 210-347 <TSA>
A:cross-references: GB:M5093; NID:918225; PIDD:RA559612.1; PJD:818225
R:king: P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,
J. Immunol. 142, 3503-3509, 1989
A:title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A:reference number: 156024, MUID:89233554; PMID:2715633
A:accession: 156024
A:status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: mRNA
A:residues: 182-201 <RS>
A:cross-references: GB:M26316; NID:951662; PIDD:AAA36336.1; PJD:9516623
R:andler, S.G.B. 733-737, 1990
A:title: Elements of secondary structure in a human epithelial mucin core peptide fragm
A:reference number: S09706, MUID:90253387; PMID:2339983
A:accession: S09706
A:molecule type: protein
A:residues: 182-201 <TEN>
C:genetics:
A:gene: GDB:MUC1, PUM
A:cross-references: GDB:120705; OMIM:158340
A:map position: 1q21-q23
C:keywords: alternative splicing, tandem repeat
F:1-23/Domain: signal sequence #status predicted <SIO>
F:24-34/Product: mucn 1, secreted epithelial tumor antigen splice form #status predict
F:24-163, 204-34/Product: mucn 1, secreted breast-cancer-associated splice form #statu
Query Match 62.4%; Score 1599; DB 2; Length 347;
Best Local Similarity 93.8%; Pred. No. 7.5e-75;
Matches 315; Conservative 0; Mismatches 1; Indels 20; Gaps 1;
QY 1 MTPGTGSPFFLLTLVTVTGSSGASSTPGGKETSATQRRSVSSSTETKNAVSTSSV 60
Db 1 MTPGTGSPFFLLTLVTVTGSSGASSTPGGKETSATQRRSVSSSTETKNAVSTSSV 60

```

Qy 61 LSSHPSGSGSTTGGQDVTLLAPATEPASGSAATWGDVTSVPTRPALSGTTPAHVTS 120
Db 61 LSSHPSGSGSTTGGQDVTLLAPATEPASGSAATWGDVTSVPTRPALSGTTPAHVTS 120
Qy 121 APPDKPP-----PGSTAPPAHGVTSAPDRPPPGSTAPPAHGVTS 160
Db 121 APPDKPP-----PGSTAPPAHGVTSAPDRPPPGSTAPPAHGVTS 160
Qy 161 APPDKPP-----PGSTAPPAHGVTSAPDRPPPGSTAPPAHGVTS 180
Db 161 APPDKPP-----PGSTAPPAHGVTSAPDRPPPGSTAPPAHGVTS 180
Qy 221 ATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 280
Db 221 ATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 280
Qy 241 ATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 300
Db 241 ATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTGV 300
Qy 281 SFFFLSHINLQFNSSLEDPSTDYQELORDISEM 316
Db 281 SFFFLSHINLQFNSSLEDPSTDYQELORDISEM 316
Qy 301 SFFFLSHINLQFNSSLEDPSTDYQELORDISEM 336
Db 301 SFFFLSHINLQFNSSLEDPSTDYQELORDISEM 336

```

RESULT 3

```

A39344
tumor-associated mucin (MUC1) homolog precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 23-Feb-1997
C/Accession: A39344
R/Spicer, A.P.; Parry, G.; Patton, S.; Gendler, S.J.
J. Biol. Chem. 266, 15099-15109, 1991
A/Title: Molecular cloning and analysis of the mouse homologue of the tumor-associated m
a loss of metastatic-like polymorphism.
A/Reference number: A39344; MUID:91332029; PMID:1714452
A/Accession: A39344
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-630 <SPI>
A/Cross-references: GB:M64928
C/Keywords: cytoskeleton; transmembrane protein

```

```

Query Match 46.9%; Score 1202.5; DB 2; Length 630;
Best Local Similarity 45.0%; Pred. No. 2.3e-54;
Matches 286; Conservative 63; Mismatches 141; Indels 145; Gaps 14;

```

```

Qy 1 MTPGTOSPFLLLLTLV-----TVTGSCHAST-----PG 32
Db 1 MTPGIRAPFLLLLSLKGFALPSEBNSVTSQDTSSSLASTTPVHSSNDPATRPP 60
Qy 33 GEKETSATORSVPSSTTEKNAVSMTSVLSHS-----PGSGSTTGGQDVTLLAPATEP 86
Db 61 GDSSTSPVQSTSSPATRAPEDSTTAVALSGTSSPATRAPVNSASSPVAGHDTSSPATSP 120
Qy 87 A--SGSAATWGDVTSVPTRPALSGTTPPAHGVTSAPDRPPPGSTAPPAHGVTSAPD 144
Db 121 LKDSNSSPVHSGTSSPATRAPVDSSTPVHGGTSSPATSPGDSSTSPHSGTSSPAT 180
Qy 145 RPPPGST-----APAAHGVTSAP 162
Db 181 RAPEDSSTTAVALSGTSSPATRAPVDSSTSPVAHDTSSPATSLSDSASSPVAGHGTSS 240
Qy 163 DTRPAFGSTAPPAH-----VTSAPDNR-----PALGSTAPPAHNVTS-- 200
Db 241 PATSPLRDSTSPVHSSASIONIKTSDLASTPDHNGSVTTSSALGATSPDHSGTST 300
Qy 201 -----ASGSA-----SGSASTLVHNGTSA 220
Db 301 TNSGSVLAATTPVYSSMPSTTKVTSGSAIIPDHNGSVLPTSSVLAGATSLVYN-TSAI 359
Qy 221 ATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTG 280
Db 360 A-TTVNSGTPSPSPQVTSVPTMAATSSHSTTASSSYSTVPTFTSSNS-SFQSLVGV 417
Qy 281 SFFFLSHINLQFNSSLEDPSTDYQELORDISEMFLQIYKQGFGLSNIKFRPGSV 340

```

```

Db 418 SFFFLFFFIQHHPNSSLEDPSSNYYQELKXNISGLEIF-NGPFLDISIKRSGSVV 476
Qy 341 VOLTLAPREGTINHVETQFNQYKTEAASRYLITSDVSHPVFPFSAQSGAGVPGNG 400
Db 477 VESTVVFREGTFSASDVKSQLIQHKKEADS-YNLTISEVKVNMQFPFSAQSRPGVGMG 535
Qy 401 IALLVLCVVALATVYLLALAVCCCRKQNGOLDIFPARDTYHMSSEPTTYHGRVYP 460
Db 536 IALLVLCVVALATVYLLALAVCCCRKKSQGLDIFPDQTYHMSSEPTTYHGRVYP 595
Qy 461 PSTDRSPYEVNSAGNGSGSLSTTPPAVTAATSA 495
Db 596 PSTDRQPYEVNSAGNGSSLSLSTTPPAVTTSA 630

```

RESULT 4

```

152257
epistatin - mouse
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C/Accession: 152257; 165210
R/Vos, H.L.; De Vries, Y.; Hilkens, J.
Biochem. Biophys. Res. Commun. 181, 121-130, 1991
A/Title: The mouse epistatin (Muc1) gene and its promoter. Rapid evolution of the repet
A/Reference number: 152257; MUID:92068178; PMID:1958179
A/Accession: 152257
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-631 <RBS>
A/Cross-references: GB:M77226; NID:g199835; PIDN:AAA39754.1; PID:g199837
A/Accession: 165210
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-631 <R2>
A/Cross-references: GB:M84683; NID:g199842; PIDN:AAA39756.1; PID:g199843
C/Genes:
A/Genes: Muc1
A/Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

```

```

Query Match 46.9%; Score 1201; DB 2; Length 631;
Best Local Similarity 45.3%; Pred. No. 2.7e-54;
Matches 288; Conservative 63; Mismatches 139; Indels 146; Gaps 15;

```

```

Qy 1 MTPGTOSPFLLLLTLV-----TVTGSCHAST-----PG 32
Db 1 MTPGIRAPFLLLLSLKGFALPSEBNSVTSQDTSSSLASTTPVHSSNDPATRPP 60
Qy 33 GEKETSATORSVPSSTTEKNAVSMTSVLSHS-----PGSGSTTGGQDVTLLAPATEP 86
Db 61 GDSSTSPVQSTSSPATRAPEDSTTAVALSGTSSPATRAPVNSASSPVAGHDTSSPATSL 120
Qy 87 A--SGSAATWGDVTSVPTRPALSGTTPPAHGVTSAPDRPPPGSTAPPAHGVTSAPD 143
Db 121 SKDSNSSPVHSGTSSPATRAPVDSSTPVHGGTSSPATSPGDSSTSPHSGTSSPAT 180
Qy 144 TRPPGST-----APAAHGVTSAP 161
Db 181 TRAPEDSSTTAVALSGTSSPATRAPVDSSTSPVAHDTSSPATSLSDSASSPVAGHGTSS 240
Qy 162 DTRPAFGSTAPPAH-----VTSAPDNR-----PALGSTAPPAHNVTS-- 200
Db 241 PATSPLRDSTSPVHSSASIONIKTSDLASTPDHNGSVTTSSALGATSPDHSGTST 300
Qy 201 -----ASGSA-----SGSASTLVHNGTSA 219
Db 301 TNSGSVLAATTPVYSSMPSTTKVTSGSAIIPDHNGSVLPTSSVLAGATSLVYN-TSAI 359
Qy 220 RATTPPAKSTPSPSPSHSDPTTLLASHSTKTDSSTHSTVPPPLTSSNHSSTPOLSTG 279
Db 360 IA-TTVNSGTPSPSPQVTSVPTMAATSSHSTTASSSYSTVPTFTSSNS-SFQSLVGV 417
Qy 280 VSEFLSHINLQFNSSLEDPSTDYQELORDISEMFLQIYKQGFGLSNIKFRPGSV 339

```



```

Query Match      10.4%; Score 265.5; DB 2; Length 1334;
Best Local Similarity 33.7%; Pred. No. 2.9e-06;
Matches 91; Conservative 14; Mismatches 90; Indels 75; Gaps 12;

Qy 31 PGEKETATQSSVPSSTKNAVMTSSVLSSHPGSGSSTGGQDVTLPATPEPASGS 90
Db 237 PGEELATTIAELLS-PSPPTPGSRTPGMTSGPCGASGAAAGSTDV-----ASGA 287

Qy 91 AATWGDVTSVTPTRPAL-----GSTTPADVTSAIDNKPAPG- 129
Db 288 GAASGDPASGASGAPVAPGSGGAPGWPAPGTAPGSGSTAPPHDTASADTAPAPGP 347

Qy 130 ----STAPAHGVTSAPDTR-PPGS-----TAPAHGVTSAPDTRPAGSTAP- 173
Db 348 TSAPGAPAHGAPAHGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 406

Qy 174 ----PAHGVTSAPDNRPALGSTAPVHNVTASGSA-----SGSASTLVH 214
Db 407 AAGTAPVAGTTPAPGTAAPAPGST--PAPGTVAPGTAAPGAPGAPDGRPPVTPGPA 460

Qy 215 NGTSATATTP--ASKSTPSIFSHTSDTP 242
Db 461 -GTGPGATPPEAAAAASASGASAPAPGEP 489

RESULT 8
A48018
mucin 7 precursor, salivary - human
N:Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
C:Species: Homo sapiens (man)
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
C:Accession: A48018; #S29115; S29116; S29114
J:Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
R:Biochem. J. 287, 639-643, 1992
A:Title: Molecular cloning, sequence, and specificity of expression of the gene encoding
A:Reference number: A48018; PMID:9338636; PMID:7690757
A:Accession: A48018
A:Molecule type: mRNA
A:Residues: 1-377 <BDB>
A:Cross-references: GB:L13283
A:Experimental source: submandibular gland
A>Note: sequence extracted from NCBI backbone (NCBIN:137719, NCBI:P:137720)
R:Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
Biochem. J. 287, 639-643, 1992
A:Title: Structural features of the low-molecular-mass human salivary mucin.
A:Reference number: S29114; PMID:9307506; PMID:1445223
A:Accession: S29115
A:Molecule type: mRNA
A:Residues: 143-168 <RED>
A:Accession: S29116
A:Molecule type: protein
A:Residues: S',71-79,'N',81-86,'XX',89,'X',91,'P' <RE2>
A:Molecule type: protein
A:Residues: 143-145,'X',147,'XXX',151-152,'X',154-158,'X',160-161,'A',163-164,'X',167-1
C:Genetics:
A:Gene: GDB:MUC7
A:Cross-references: GDB:138799; OMIM:158375
A:Map position: 4q33-4q31
A:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:13-377/Product: mucin 7, salivary #status predicted <AMT>
F:97,128,135,146,312/Binding site: carbohydrate (Aam) (covalent) #status predicted

Query Match      9.4%; Score 240; DB 2; Length 377;
Best Local Similarity 30.5%; Pred. No. 1.5e-05;
Matches 90; Conservative 44; Mismatches 91; Indels 70; Gaps 17;

Qy 37 TSATQSSVPSSTKNAVMTSSVLSSHPGSGSSTGGQDVTLPATPEPASGSAATGQ 96
Db 118 SASRTITLPLNT--FLPQWATITSSRENVMTSSV--ATLAPVNSAP-----Q 163

Qy 97 DVTSVTPRALGSTTPADVTSAIDNKPAPGSR--APAHGVTSAPDTRPPGSTAPA 154

```

```

Db 164 DTTAAP--PTPSATTP-----APSSAPPETTAAPP-----TPSATPAPSSAAP 209
Qy 155 AHGVTSAPDTRPA-----PGSTAPAHGVTSAPDNRPA-----LGTAPVHNVTASAGS 204
Db 210 E--TTAAPPTPPTTAPPSAPPE--TTAAPPTSATTPAPLSSAPF-----ETTAAP 261
Qy 205 ASGSASTLVHNGTSARATTT-----PASKSTPSIFSHTSDT-----PTTLASH 248
Db 262 PTPSATTLDPSSASAPPETTAPPTPSATTPAPPSAPAPERTTAPITTPSSPTTLAPD 321
Qy 249 STKTDASTSTHSTVPLTSSNHS13POLSTG--VGFPLSPHINLOPNSLED 300
Db 322 TSEISAPHTOTTTSVTTOTTTTTPKOTSAPOKNIKSRFL--YMKL-LNKIID 373

RESULT 9
T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Geisel, C.; Gattling, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AACT0889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 9'

Query Match      9.3%; Score 237; DB 2; Length 1032;
Best Local Similarity 30.5%; Pred. No. 6.2e-05;
Matches 103; Conservative 42; Mismatches 131; Indels 62; Gaps 14;

Qy 16 TVLVYVVGSG--HASTP-----GGEKETATQSSVPSSTKNAVMTSSVLSSHPG 67
Db 357 GTVTVVVGSSSTPASTPTIASSSSGSTVTVVAPGSSSTVSGSSTPSSSGSTWSTVSGS 416
Qy 68 SSSSTTGGQDVTLPATPEPASG-----SAATWGDVTSVTPTRPALGSTTPPAHDVTS 121
Db 417 TGST-----VTAPVSSSTPSSSTPPLASSSSSGSTVTVVSSSGSTTPASSSSAG 470
Qy 122 PDKRPA-----PGSTAPAHGVTSAPDTRPPGST--APAHGVTSAP--DTRPA 167
Db 471 TASTISGSGSTATTIVPGSSSVGSSVQSASPSFPTWSTVSGPTGSTTVVPGSSTP 530
Qy 168 PGSTAPAHGVTSAPDNRPALGSTAPVHNV--TSASGASGASTLVHNGTSARATTP 225
Db 531 PSSSNPSSSPASTGSTTITISGSSGSIIVTVSGSTVSGSTVSGSTL-----ASSTAP 584
Qy 226 ASKSTPSI-----PSHSDPTPTLASHTKTDASTHSTVPL--TSSNHS----- 272
Db 585 GSSSTVSSSSPPSSQSPAPVTSSTTPSGTSSQSPSPMNSSTTPGSGSGSTTPSS 644
Qy 273 ---SPQLSTGVSPFLSFH---SLQPNLSLEDPESTD 304
Db 645 TMSPTGSGT-STFSVATEVTSQSTVPSGSLGTOSTN 681

RESULT 10
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattling, S.

```

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid K06A9.

A:Reference number: Z21525

A:Accession: T34434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2332 <GB>

A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a

A:Genetic source: strain Bristol N2; clone K06A9

A:Gene: CESP:K06A9.1a

A:Map position: X

A:Insertion: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 9.3%; Score 237; DB 2; Length 2232;

Best Local Similarity 30.5%; Pred. No. 0.00014;

Matches 103; Conservative 42; Mismatches 131; Indels 62; Gaps 14;

16 TULVTVTGSG--HASSTP-----GGEKETSATORSSVPSSTERNVAMTSVLSHSPG 67

357 SYTVVVGSSSTRASSTPLASSSPGSTVTVAPSSSTGSSSTPSSASSSGTMTNSGS 416

68 SGGSTTGGQDVTAPRTEPAG-----SAATGQDVTVPYTRALGCTTPPAHDVTS 121

417 TGGT-----VIVAPVSSSTFGSSSTPLASSSSSGSTVTVVSSSSSTPSSASSSG 470

122 PDKPA-----PGSTAPPAHGVTSAPDTRPPGST--APAAHGVTSAP--DTRPA 167

471 TASTISGSGSTATTATVPGSSSSVSSSTQSSAPSSPQMTGTVVVGSSSTSPA 530

168 PGSTAPPAHGVTSAPDTRPPGST--APAAHGVTSAP--DTRPA 167

531 PSSSPNSSSPASTGTTTSSSSIIIVSGSTVSGSTGTSQTL-----ASTATP 584

226 ASKTPST-----PETHSDPTTLASHTKTDASSTHHSTVPL--TSNHSHT----- 272

585 GSSSTVPSSSPQSPAPVTSITPSSQTSQSPSPMNSSTPTGSSQSTTPESG 644

273 ---SPQSTGVSPFLSPFI---SNLQFNSSLEDPST 304

645 TASSPTGSGT-STFSVATEVTSOSTVPSGSLGTOSTN 681

RESULT 11

OOBZ1

membrane antigen gp350 - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4; Epstein-Barr virus

C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 28-Jul-2000

C:Accession: A43042; S33008; S33009; A03762

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A:Reference number: A93065; MUID:85035713; PMID:6092825

A:Accession: A43042

A:Molecule type: DNA

A:Residues: 1-907 <BAN>

A:Cross-references: GB:V01555; GB:J02070; GB:K01729; GB:K01730; GB:V01554; GB:X00498; GB

R:Farrell, P.J.

submitted to the EMBL Data Library, March 1988

A:Reference number: S32973

A:Accession: S33008

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 <FAR>

A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24855.1; PID:g1334869

A:Accession: S33009

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-907 <FA2>

A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24855.1; PID:g1334869

R:Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667; PMID:6087149

A:Contents: annotation; protein coding region

C:Superfamily: Epstein-Barr virus membrane antigen gp350

Query Match 9.2%; Score 236.5; DB 1; Length 907;

Best Local Similarity 23.1%; Pred. No. 5.8e-05;

Matches 123; Conservative 62; Mismatches 194; Indels 153; Gaps 22;

22 TSGGHAASST-----PGGEKETSATORSSVPSSTERNVAMTSVLSHSPGSSST 73

444 TTGGLPSSTHVPNTLTPASTGPTVSF--ADVTSPTPAGTTSGASVTPSPMDNGTES 501

74 QGQDVT--LAPATEPASGSAATWGQDVTSV-----VTRPALGCTTPPAH 116

502 KADDMSTSTSPVTPPTNATSPPTAVTTPPTNATSPPTAVTTPPTNATSPPTGKTSPTSA 561

117 DVTSAPD--NKPAPGSTAPPAHGVTSAPD--TRPPGSTAPPAHGVTSAPD--TRPAGSTAP 173

562 VTTPTPNTATSPPTGKTSPTSAVTTPTPNTATSPPTGKTSPTSAVTTPTPNTATSPPTGKTSPT 621

174 PAHG-----VTSAPDNRPALGCTTPPAHGVTSAGSA-----SGSASTL-- 212

622 QANANHTLGTSPVPTVTSQPKNATSAVTTGQ--HNTSSSTSSMLRPSNPPTLSPS 679

213 -VANGTSARATTPA-----SKSTPPSIPSHSDPTTLASHTKTDASTHNSHT-- 261

680 TSDNMTSHMPLTSAHPTGEMITQVTPASISITHVSTSSPAPRPGTTSQASGPNSSSTS 739

262 -----VPELTSNHSSTPOLSTGVSPFLSPHISNLOFNSSL-----BDP 301

740 TKRGEVNTKATPPQNNATSPQASGQKTAVPVTVSGKXNSTGKXHTTGKARTSTEP 799

302 STDYVQELQORDISEMFLQITKQGFGLSNIKFRGSAVVQUTLARBEQTNVHDVETQF 361

800 TTDY-----GGDSTTPRPYNTATYLPST----- 824

362 NQYKTEASRYNLTISDVSYH--VPPPSAQSAGVPGW-GIALVLV-CVLVALAVY 417

825 ---SSKLRPRRTFTSPVTTAQTAVVPPTSQ---PRFSNLSMVLQOMASIALVLTLL 876

418 LIALAVCCRRKNTGQLDIFPARDVHPMSEKPTVTHGRVYVPSSTDSPPV 469

877 LVMADCAFR-----NLSHTSHT---YTPPYDDAEFY 906

RESULT 12

JH0557

exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi

N:Alternate names: neuraminidase

C:Species: Trypanosoma cruzi

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000

C:Accession: JH0557

R:Perreira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.

J. Exp. Med. 174, 179-191, 1991

A:Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial nei

A:Reference number: JH0557; MUID:91277609; PMID:1711561

A:Accession: JH0557

A:Molecule type: DNA

A:Residues: 1-1162 <PER>

A:Cross-references: GB:M61732; NID:g162302; PID:g162303

A:Note: the authors translated the codon TCT for residue 45 as Cys

C:Comment: This protein plays a role in parasite-host cell interaction.

C:Superfamily: trypanastigote-specific surface antigen

C:Keywords: glycoprotein; glycosidase; hydrolase

F/394/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 9.1%; Score 234; DB 2; Length 1162;

Best Local Similarity 28.7%; Pred. No. 0.0001;

Matches 87; Conservative 34; Mismatches 134; Indels 48; Gaps 11;

16 TULVTVTGSG-----ASSTPGGEKETSATORSSVPSSTERNVAMTSVLSHSPG 65

707 TSTVDSASAGTSPTRPADSASHTSPTRPADSASHTSPTRPADSASHTSPTRPADSASHTSPTRPADSASHT 766

[illegible]

RESULT 13
548478
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
N/Alternative names: extracellular glucanase; mucin-like protein MUC1; protein YIR019c
C/Species: *Saccharomyces cerevisiae*
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C/Accession: S48478; A26877; B26877; S27281; J06123
R/Rowley, K.
submitted to the EMBL Data Library, October 1994
A/Reference number: S48478
A/Accession: S48478
A/Molecule type: DNA
A/Residues: 1-1367 <ROM>
A/Cross-references: GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GSPDE:GNO0009; MIF
J.Yamashita, I.; Nakamura, M.; Fukui, S.
U. Bacteriol. 169, 2142-2149, 1987
A/Title: Gene fusion is a possible mechanism underlying the evolution of STR1.
A/Reference number: A91831; MUID:87194600; PMID:3106330
A/Accession: A26877
A/Molecule type: DNA
A/Residues: 1-242 <YAM>
A/Cross-references: EMBL:M16164; NID:G172522; PIDN:AAA35014.1; PID:G172525
A/Accession: B26877
A/Molecule type: DNA
A/Residues: 762-1331 <YAZ>
A/Cross-references: EMBL:M16165; NID:G172523; PIDN:AAA35015.1; PID:G172526
R.Pardo, J.M.; Jance, E.; Zalacain, M.; Clatos, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A/Title: Similar short elements in the 5' regions of the STR2 and SGA genes from *Sacchar*
A/Reference number: S27281; MUID:89031230; PMID:3141213
A/Accession: S27281
A/Molecule type: DNA
A/Residues: 1-31 <PAR>
A/Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:G4552
R.Lambrecht, M.G.; Bauer, F.F.; Marmur, U.; Peticola, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A/Title: Muc1, a mucin-like protein that is regulated by Msa10, is critical for pseudoph
A/Reference number: J06123; MUID:9632337; PMID:8710886
A/Accession: J06123
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1367 <LAM>
A/Cross-references: GB:030626; NID:G1304386; PIDN:AA49609.1; PID:G1304387
C/Genetics:
A/Gene: SGD:MUC1; STR2; MAL5; DEX2; SGD:S0001458
A/Cross-references: MIPS:YIR019c; SGD:S0001458
A/Map position: 9R
C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C/Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
P:5-21/Domain: transmembrane #status predicted <TM1>
P:1350-1366/Domain: transmembrane #status predicted <TM2>

| | | | | |
|------------------|-------|-------------------|------|------------------------------------|
| Query Match | 9.1% | Score 233.5 | DB 1 | Length 1367 |
| Local Similarity | 28.9% | Pred. No. 0.00013 | | |
| Matches | 88 | Conservative | 45 | Mismatches 123; Indels 49; Gaps 12 |

| | | | | | |
|----|-----|---|----------------------|----------|-----|
| QY | 20 | VVTSGGASHSTPGGEKETSATORSSV--- | SSTEKNVAMTSSVLSSHS-- | PGSGSSTT | 73 |
| | | | | | |
| Db | 344 | VTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTPSSSTT | | | 403 |
| | | | | | |
| QY | 74 | QGDVTLATPEPSSGAATFWG-QDVTVPTLRALGSTTPRHADVTSAIPDNKAP--- | | | 128 |
| | | | | | |
| Db | 404 | ESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTSS | | | 463 |
| | | | | | |
| QY | 129 | -----SSTAPRHAGT-----SAPDTRPPGTSIPAAHGVTSAPDTR----- | PAPGSTAPPAH | | 176 |
| | | | | | |
| Db | 464 | STTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTES- | | | 522 |
| | | | | | |
| QY | 177 | GVTSAPDNRPALGST--APVHNVTASGSA--SGSASTLVHNGTSARATTTPAKS- | | | 229 |
| | | | | | |
| Db | 523 | --SSAPAPPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSTTPVTSSTTESSSAP | | | 580 |
| | | | | | |
| QY | 230 | --TPSISIHSDPTFLASHST-----KTASSTHSTVPLTSSNHS- | | | 273 |
| | | | | | |
| Db | 581 | VPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPT | | | 640 |
| | | | | | |
| QY | 274 | POLST | | | 278 |
| | | | | | |
| Db | 641 | PSST | | | 645 |

```

RESULT 14
T21460
hypothetical protein ZK945.10 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
R|Accession: T21460
R|Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A|Reference number: Z19425
A|Accession: T21460
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-796 <WLI>
A|Cross-references: EMBL:Z48582; PIDD:CAA8469.1; GSPDB:GN00020; CESP:ZK945.10
A|Experimental source: clone F27B5
C|Genetics:
A|Gene: CESP:ZK945.10
A|Map position: 2
A|Introns: 62/3; 124/3; 169/1; 216/1; 262/3; 371/1; 581/3; 608/1

Query Match          9.1%; Score 233; DB 2; Length 796;
Best Local Similarity 26.8%; Pred. No. 7.5e-05;
Matches 115; Conservative 54; Mismatches 172; Indels 88; Gaps 19

OY      19 TVVTGSGHASTPGGE---KEISAIQRGSVPSETEKNVAMTSVLISHSPGSG-----70
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       302 TVTTAMSTSTPSTPTSTTISTSTPTTSTASTASTSASTTGQSSSTITSSPESTLTSTSI 361
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY      71 STTGQDVTLPATEPASGSAATMGQDVSVVTRPALGSTTPPADVTSADNKPAPGS 130
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       362 PTTFPELTSTLSIPDNALCSLYDETTSTFTFMTLTSTP-----TEEP-----S 408
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY      131 TAPPAGVTSAPDRPPPGSTAAMAAGVTSAPDRRAPAGSTAAPPAIGVTSAPDNRPALGS 190
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       409 TSPTTTEVTSTST--VTTEPTTTLTSTAST-----STTESSTSTVTTSP-----S 454
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY      191 TAPPVNVTVASGSGASGASGLTNHTSARAT-TTPAS----KSTPSI-----PSGHSDT 241
           |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       455 TSPVSTVTVSSG-----SSSTVTPTPTSTESTSTSPSIVTSTTAPASTSTGGPSSSST 509
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY      242 PTLASHSKTKDASTAHST--VPLTGSNNHSNGPOLSGVGFPL-----SPITS---290
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db       510 PSSSTASSVSSTASTSGSTSTQQSSTTKSHTTSSDGNNPDPIVEAKATTTFIDSTSV 569
           :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

QY 291 NLOFNSL-----EDPSTDYQELQORDISMEFLQ-----IKQGGFLGJS 330
Db 570 NLTNLSGLIIGVQTSICTSPSTSNVSTTKD-GACFKSVSMRLGITYPASTFVGP 628
QY 331 NIKFRGSAV-----VOLILA--FREGTINHDVETQFQVTEAASRYNLTIDVSVSH 383
Db 629 NYTFRATMTTDKAKYYIYANVYIQEYSSTTIESBSTSAVASSTSPSPSTLSTST 688
QY 384 VPPFPAQS 392
Db 689 VTEPSTRS 697

RESULT 15

T45025
mucin MUC5B, tracheobronchial (imported) - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text change 21-Jul-2000
C/Accession: T45025
R/Deeseyn, J.L.; Guyonnet-Dupercat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A/Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A/Reference number: Z22899; MUID:9716151; PMID:9013550
A/Accession: T45025
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3570 <DESeq>
A/Cross-references: EMBL:Z22496; NID:g1834502; PID:CA96577.1; PID:g1834503
A/Experimental source: placenta
C/Genetics:
A/Gene: MUC5B

Query Match 9.1%; Score 233; DB 2; Length 3570;
Best Local Similarity 23.6%; Pred. No. 0.00038;
Matches 160; Conservative 50; Mismatches 215; Indels 252; Gaps 31;

QY 2 TPQTQSPFLILLTLVTVVTS---GHASSTPGGR---KETSATQSSVP-----SS 48
Db 1775 TPQTT---WILTELTATTTAALPHGTSPSTPGTWTILTEPSTTATVTVTGSTATASS 1831
QY 49 TEKNA-----VMTSSVLSH-SFGSSSTTQGDVTLAPATEPASGSAATWGQDT 99
Db 1832 TRATAGTLKVLSTVATPTVSSRATPSSPGATAPALRSTATTPTATS-----VT 1884
QY 100 SVPTRPALG-----STTPADVTSAADNKP-----APG 129
Db 1885 AIRPS--SLGTWTRLSGTTPPTATWSTATPSTPETVHTSVLTTATTTTSGVATPS 1942
QY 130 STAPRAH-----GVTSAPDTRP-----P-----GSTPAAHGVTSAPDTRP----- 166
Db 1943 STPGIATHTTKVPTTTTTFIFATPSSPGIATTPVWISTTTPTTGTGTVTPSSIPGTH 2002
QY 167 -----APGSTAP-----AHGVTSAPDNRPALGST-A 192
Db 2003 TATVLTATTTTVAATGSMATPSSSTQSGTPPSLTATTATTTATGSTMPSSTP--GTTPI 2060
QY 193 PPHANTSSAGSAGSAST-----LVHNTSARATTPASKSTPSP-- 235
Db 2061 PPVLTATTPATSSITVTPSSALGTTTPPVNTATTHGSLPSSPHTVPTAMTSAT 2120
QY 236 -----SHSDPTTTLASHSTKTDASTHSTVPLTSSNH-----STSPOLS 277
Db 2121 SGILGTHITEPSTGTHSTPATGTTQPSST--PALSPHSRTSPSPGTTTGH 2178
QY 278 TGVSFFEL-----PHINLOFNSLEDPTDYQ--ELQORDISEMFLQIYKO-----G 324
Db 2179 RGTSTRTATATPSKRTISTILPSSPTAPITTVTTGCPQCAWSEMLDYSYMPGSPSG 2238
QY 325 GFGLSNIKFRGSAVVO-LTLAFREGTINHDVETQFQVTEAASRYNLTIDVSVSH 383
Db 2239 DFTYTSNIRAAAGAVCEQPLGLECR----- 2263

QY 384 VPPFPAQSGAGVPGKGIALLV-----LYC-----VLVALATVLIATAVQCCKRN 430
Db 2264 -----AQAPGVPLRELGQVVECSLDPLGLVCNRBQVSKFMCKNVEIRVFC-----N 2312
QY 431 YGOLDIPARDTYHMPSEYP-----TYHTGRVYVPPSTDSPPEKVSAGN 476
Db 2313 YGHCPSRPAATSSATPSSSTPGTWTILKLTATTATTSTSTGSTATPSSSQGFP-----AGT 2367
QY 477 GGSSLSTYTPAVALTSA 493
Db 2368 PHVSTATPTPTVTSKA 2384

Search completed: May 1, 2003, 23:13:51
Job time : 60.0536 secs

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.rat

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:21:52 ; Search time 28 Seconds
(without alignments)
9.457 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMPLOIY 9

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: * 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 66399

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum March 0%
Maximum March 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgm2_6/prodata/1/iaa/5A COMB pep:*
2: /cgm2_6/prodata/1/iaa/5B COMB pep:*
3: /cgm2_6/prodata/1/iaa/6A COMB pep:*
4: /cgm2_6/prodata/1/iaa/6B COMB pep:*
5: /cgm2_6/prodata/1/iaa/6C COMB pep:*
6: /cgm2_6/prodata/1/iaa/6D COMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---|
| 1 | 21 | 47.7 | 9 | 1 | US-08-338-634-10 Sequence 10, Appl |
| 2 | 21 | 47.7 | 9 | 3 | US-08-339-339A-218 Sequence 25, Appl |
| 3 | 21 | 47.7 | 9 | 5 | PCT-US95-16415-25 Sequence 25, Appl |
| 4 | 20 | 45.5 | 5 | 4 | US-08-364-121-9 Sequence 3, Appl |
| 5 | 20 | 45.5 | 5 | 4 | US-08-364-121-9 Sequence 3, Appl |
| 6 | 20 | 45.5 | 9 | 3 | US-08-459-339A-336 Sequence 36, App |
| 7 | 20 | 45.5 | 9 | 3 | US-09-227-337-660 Sequence 80, App |
| 8 | 19 | 43.2 | 8 | 3 | US-09-121-964-5 Sequence 3, Appl |
| 9 | 19 | 43.2 | 9 | 1 | US-08-367-758B-8 Sequence 8, Appl |
| 10 | 19 | 43.2 | 9 | 2 | US-08-367-758B-8 Sequence 8, Appl |
| 11 | 18 | 40.9 | 6 | 4 | US-08-179-558-52 Sequence 52, Appl |
| 12 | 18 | 40.9 | 7 | 1 | US-07-678-492-13 Sequence 13, Appl |
| 13 | 18 | 40.9 | 7 | 3 | US-08-556-418-14 Sequence 4, Appl |
| 14 | 18 | 40.9 | 7 | 4 | US-08-747-599A-4 Patent No. 5240848 |
| 15 | 18 | 40.9 | 7 | 6 | 5420848-3 Patent No. 5422249 |
| 16 | 18 | 40.9 | 9 | 1 | US-08-338-634-12 Sequence 12, Appl |
| 17 | 18 | 40.9 | 9 | 1 | US-08-338-634-25 Sequence 30, Appl |
| 18 | 18 | 40.9 | 9 | 1 | US-08-338-634-30 Sequence 30, Appl |
| 19 | 18 | 40.9 | 9 | 4 | US-08-747-599A-29 Sequence 29, Appl |
| 20 | 18 | 40.9 | 9 | 4 | US-08-747-599A-29 Sequence 29, Appl |
| 21 | 18 | 40.9 | 9 | 4 | US-09-187-859-3138 Sequence 3138, Appl |
| 22 | 18 | 40.9 | 9 | 6 | PCT-US95-16415-35 Sequence 35, Appl |
| 23 | 17 | 38.6 | 6 | 5 | 5200320-15 Patent No. 5200320 |
| 24 | 17 | 38.6 | 7 | 1 | US-08-180-209B-8 Sequence 8, Appl |
| 25 | 17 | 38.6 | 7 | 1 | US-08-385-745-8 Sequence 8, Appl |
| 26 | 17 | 38.6 | 7 | 1 | US-08-222-619-17 Sequence 17, Appl |
| 27 | 17 | 38.6 | 7 | 4 | US-08-747-599A-26 Sequence 26, Appl |

| | | | | | |
|----|----|------|---|---|---|
| 28 | 17 | 38.6 | 7 | 4 | US-08-485-389-8 Sequence 8, Appl |
| 29 | 17 | 38.6 | 7 | 4 | US-08-474-853-8 Sequence 8, Appl |
| 30 | 17 | 38.6 | 7 | 4 | US-09-166-205B-8 Sequence 8, Appl |
| 31 | 17 | 38.6 | 7 | 5 | PCT-US94-02629-8 Sequence 17, Appl |
| 32 | 17 | 38.6 | 9 | 1 | US-08-340-283-181 Sequence 181, App |
| 33 | 17 | 38.6 | 9 | 2 | US-08-340-283-181 Sequence 321, App |
| 34 | 17 | 38.6 | 9 | 2 | US-08-146-028-320 Sequence 321, App |
| 35 | 17 | 38.6 | 9 | 2 | US-08-146-028-321 Sequence 321, App |
| 36 | 17 | 38.6 | 9 | 2 | US-08-146-028-322 Sequence 321, App |
| 37 | 17 | 38.6 | 9 | 2 | US-08-146-028-327 Sequence 327, App |
| 38 | 17 | 38.6 | 9 | 2 | US-08-146-028-328 Sequence 328, App |
| 39 | 17 | 38.6 | 9 | 2 | US-08-146-028-328 Sequence 328, App |
| 40 | 17 | 38.6 | 9 | 2 | US-08-146-028-334 Sequence 334, App |
| 41 | 17 | 38.6 | 9 | 4 | US-08-723-425A-320 Sequence 320, App |
| 42 | 17 | 38.6 | 9 | 4 | US-08-723-425A-321 Sequence 321, App |
| 43 | 17 | 38.6 | 9 | 4 | US-08-723-425A-326 Sequence 326, App |
| 44 | 17 | 38.6 | 9 | 4 | US-08-723-425A-327 Sequence 327, App |
| 45 | 17 | 38.6 | 9 | 4 | US-08-723-425A-328 Sequence 328, App |

ALIGNMENTS

RESULT 1
US-08-338-634-10
; Sequence 10, Application US/08338634
; Patent No. 5679641
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Peptides of human p53 protein for use
; TITLE OF INVENTION: in human T cell response inducing compositions, and
; TITLE OF INVENTION: human p53 protein-specific cytotoxic T-lymphocytes.
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338, 634
; FILING DATE: 06-February-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL93/00102
; FILING DATE: 18-May-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-26
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3582
; TELEFAX: (516) 822-3550
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULAR TYPE: Peptide
; HYPOTHEICAL: NO
; US-08-338-634-10
Query Match 47.7%, Score 21; DB 1; Length 9;
Best Local Similarity 37.5%; Fred. No. 1.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0;
Gy 1 ISEMPLOIY 8

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.ra1

Page 2

Db 2 LINKFCOL 9

```

;:::|:
;
RESULT 2
US-08-159-339A-618
; Sequence 618, Application US/08159339A
; Patent No. 6037135
;
GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Etebeah
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
;
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;
OPERATING SYSTEM: DOS
;
SOFTWARE: PASTESQ for Windows Version 2.0
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
;
ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
;
INFORMATION FOR SEQ ID NO: 618:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
MOLECULE TYPE: peptide
;
US-08-159-339A-618
;
Query Match 47.7%; Score 21; DB 3; Length 9;
Best Local Similarity 42.9%; Pred. No. 1.9e+05;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
;
CY 3 EMFLQIT 9
;:::|:
;
Db 3 EYFVEMT 9
;:::|:
;
RESULT 3
PCT-US95-16415-25
; Sequence 25, Application PCT/US9516415
;
GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; TITLE OF INVENTION: IN VIVO ACTIVATION OF TUMOR-SPECIFIC
; TITLE OF INVENTION: CYTOTOXIC T CELLS
; NUMBER OF SEQUENCES: 38
;

```

```

;
CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10666 North Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
OPERATING SYSTEM: PC-DOS/MS-DOS
;
SOFTWARE: Patent Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16415
; FILING DATE: 13-DEC-1995
; CLASSIFICATION:
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/355,558
; FILING DATE: 14-DEC-1994
;
ATTORNEY/AGENT INFORMATION:
; NAME: Logan, Amy L. C.
; REGISTRATION NUMBER: 13,950
; REFERENCE/DOCKET NUMBER: 433,1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
;
INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
MOLECULE TYPE: peptide
;
PCT-US95-16415-25
;
Query Match 47.7%; Score 21; DB 5; Length 9;
Best Local Similarity 37.5%; Pred. No. 1.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
;
CY 1 ISEMFLQIT 8
;:::|:
;
Db 2 LINKFCOL 9
;:::|:
;
RESULT 4
US-08-964-127-9
; Sequence 9, Application US/08964127
; Patent No. 6277565
;
GENERAL INFORMATION:
; APPLICANT: Grandaearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
;
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;
OPERATING SYSTEM: Windows 95
;
SOFTWARE: PASTESQ for Windows Version 2.0b
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
;

```

REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-964-127-9

Query Match 45.5%; Score 20; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMF 5
||:|
Db 1 ISKMF 5

RESULT 5
US-09-496-692-9
Sequence 9, Application US/09496692
Patent No. 6313271
GENERAL INFORMATION:
APPLICANT: Grandearl, Andrew David John
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/496,692
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/964,127
FILING DATE: 06-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-496-692-9

Query Match 45.5%; Score 20; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMF 5
||:|
Db 1 ISKMF 5

RESULT 6
US-08-159-339A-336
Sequence 336, Application US/08159339A
Patent No. 6037135
GENERAL INFORMATION:
APPLICANT: Kubo, Ralph T.
APPLICANT: Grey, Howard M.
APPLICANT: Sette, Alessandro
APPLICANT: Celis, Esteban
TITLE OF INVENTION: HLA Binding peptides and Their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 1254
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 336:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-336

Query Match 45.5%; Score 20; DB 3; Length 9;
Best Local Similarity 22.2%; Pred. No. 1.9e+05;
Matches 2; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFQIY 9
:::|:|
Db 1 LTKQYLVIV 9

RESULT 7
US-09-227-357-660
Sequence 660, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P010P1
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684

EARLIER FILING DATE: 1998-07-07
 EARLIER APPLICATION NUMBER: 60/051,926
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,793
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,925
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,929
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,803
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,732
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,931
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,932
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,916
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,930
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,918
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,920
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,733
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,795
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,919
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,928
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/055,722
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,723
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,948
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,949
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,953
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,950
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,947
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,964
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/056,360
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,684
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,984
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/055,954
 EARLIER FILING DATE: 1997-08-18
 EARLIER APPLICATION NUMBER: 60/058,785
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,664
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,660
 EARLIER FILING DATE: 1997-09-12
 EARLIER APPLICATION NUMBER: 60/058,661
 EARLIER FILING DATE: 1997-09-12
 NUMBER OF SEQ ID NOS: 672
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 660
 LENGTH: 9
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-227-357-660

Query Match 45.5%; Score 20; DB 4; Length 9;
 Best Local Similarity 57.1%; Pred. No. 1.9e+05;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLO 7
 DB 1 IREIFLR 7

RESULT 8
 US-09-121-964-5
 Sequence 5; Application US/09121964
 Patent No. 6124447
 GENERAL INFORMATION:
 APPLICANT: Natori, Shunji
 TITLE OF INVENTION: NOVEL ENZYME CATALYZING DEPHOSPHORYLATION
 FILE REFERENCE: 32290-144753
 CURRENT APPLICATION NUMBER: US/09/121,964
 CURRENT FILING DATE: 1998-07-24
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 5
 LENGTH: 8
 TYPE: PRT
 ORGANISM: Sarcophaga peregrina
 US-09-121-964-5

Query Match 43.2%; Score 19; DB 3; Length 8;
 Best Local Similarity 50.0%; Pred. No. 1.9e+05;
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 MFLQY 9
 DB 1 IFIOY 6

RESULT 9
 US-08-367-758B-8
 Sequence 8; Application US/08367758B
 Patent No. 5767235
 GENERAL INFORMATION:
 APPLICANT: Muramatsu, Ryo
 APPLICANT: Suresada, Akiko
 APPLICANT: Misawa, Satoru
 APPLICANT: Nukui, Eiko
 APPLICANT: Wada, Koichi
 APPLICANT: Nakano, Masaharu
 APPLICANT: Morikawa, Tadamori
 TITLE OF INVENTION: Hirudin Variants, or Their Salts, Method
 TITLE OF INVENTION: For Their Production and Anticoagulants Having Said
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolach & Birch, LLP
 STREET: P.O. Box 747
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22040-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/367,758B
 FILING DATE: 03-JAN-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1516-188P

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..9
OTHER INFORMATION: /label= Example_3H_4H
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /product= "succinyl-glu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "gln-oh"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /product= "tyr or tyr-SO3"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /product= "tyr or tyr-SO3"
US-08-367-758b-8

Query Match
Best Local Similarity 57.1%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMPLO 7
Db 3 IPEYVLO 9

RESULT 10
US-08-909-735-8
Sequence 8, Application US/08909735
Patent No. 5880258
GENERAL INFORMATION:
APPLICANT: Muramatsu, Ryo
APPLICANT: Sukesada, Akiko
APPLICANT: Misawa, Satoru
APPLICANT: Nukui, Eriko
APPLICANT: Wada, Koichi
APPLICANT: Nakano, Masaharu
APPLICANT: Morikawa, Tadanori
APPLICANT: Kobashi, Kyochi
TITLE OF INVENTION: Hybrid Variants, or Their Salts, Method
TITLE OF INVENTION: for Their Production and Anticoagulants Having Said
TITLE OF INVENTION: Compounds as Active Ingredients
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSES: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,735
```

```
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/367,758
FILING DATE: 03-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1516-188P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..9
OTHER INFORMATION: /label= Example_3H_4H
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /product= "succinyl-glu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "gln-oh"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /product= "tyr or tyr-SO3"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /product= "tyr or tyr-SO3"
US-08-909-735-8

Query Match
Best Local Similarity 57.1%; Pred. No. 1.9e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMPLO 7
Db 3 IPEYVLO 9

RESULT 11
US-09-179-558-52
Sequence 52, Application US/09179558
Patent No. 6180612
GENERAL INFORMATION:
APPLICANT: Hockensmith, Joel W.
APPLICANT: Muthuswami, Rohini
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: TARGETING DNA METABOLIC PROCESSES USING
TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSES: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
```

SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/179,558
FILING DATE: 27-OCT-1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 09/060,470
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 60/063,898
FILING DATE: 31-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 9426-005-999
TELEPHONE: (212) 7909090
TELEFAX: (212) 8699741
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-179-558-52

Query Match
Best Local Similarity 40.9%; Score 18; DB 4; Length 6;
Best Local Similarity 50.0%; Pred. No. 1.9e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMFL01 8
Db 1 ELYT01 6

RESULT 12
US-07-638-498-13
Sequence 13, Application US/07638492
Patent No. 5494672
GENERAL INFORMATION:
APPLICANT: Hodges, Robert S.
APPLICANT: Paranchych, William
APPLICANT: Irvin, Randall T.
APPLICANT: Lee, Kok K.
APPLICANT: Parimi, Sashiy A.
APPLICANT: Zoultman, Dick E.
APPLICANT: Doig, Peter C.
APPLICANT: Wong, Mah Y.
TITLE OF INVENTION: Pseudomonas Peptide Composition and
METHOD OF INVENTION: Method
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,492
FILING DATE: 04-JAN-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 8900-0002.21

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: peptide 13
US-07-638-492-13

Query Match
Best Local Similarity 40.9%; Score 18; DB 1; Length 7;
Best Local Similarity 75.0%; Pred. No. 1.9e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EMFL 6
Db 2 EMFL 5

RESULT 13
US-08-556-419-14
Sequence 14, Application US/08556419C
Patent No. 6093549
GENERAL INFORMATION:
APPLICANT: Ross, Christopher
APPLICANT: Li, Xiao-Jiang
APPLICANT: Li, Shi-Hua
APPLICANT: Sharp, Alan
APPLICANT: Latham, Anthony
APPLICANT: Snyder, Paul
APPLICANT: Morley, Solomon
TITLE OF INVENTION: Huntingtin-associated protein
FILE REFERENCE: 01107, 52271
CURRENT APPLICATION NUMBER: US/08/556,419C
CURRENT FILING DATE: 1995-11-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 7
TYPE: PPT
ORGANISM: Rattus norvegicus
US-08-556-419-14

Query Match
Best Local Similarity 40.9%; Score 18; DB 3; Length 7;
Best Local Similarity 75.0%; Pred. No. 1.9e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 LQIY 9
Db 4 LQIY 7

RESULT 14
US-08-747-599A-4
Sequence 4, Application US/08747599A
Patent No. 6214795
GENERAL INFORMATION:
APPLICANT: Benjamin, Howard et al.
TITLE OF INVENTION: Peptide Compounds Useful for Modulating
PCP Receptor Activity
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

Thu May 8 16:14:07 2003

us-09-658-621b-26.closed.ra1

Page 7

ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,599A
FILING DATE: 12-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kara, Catherine J.
REGISTRATION NUMBER: 41,106
REFERENCE/DOCKET NUMBER: PPI-029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-747-599A-4

Query Match 40.9%; Score 18; DB 4; Length 7;
Best Local Similarity 60.0%; Pred. No. 1.9e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLDIV 9
Db 1 FLDIV 5

RESULT 15
5240848-3
Patent No. 5240848
APPLICANT: KECK, PAMELA J.; CONNOLLY, DANIEL T.; PEDER, JOSEPH
TITLE OF INVENTION: DNA SEQUENCES ENCODING HUMAN VASCULAR
PERMEABILITY FACTOR HAVING 189 AMINO ACIDS
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/337,037
FILING DATE: 10-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 274,061
FILING DATE: 21-NOV-1988
SEQ ID NO: 3:
LENGTH: 7
5240848-3

Query Match 40.9%; Score 18; DB 6; Length 7;
Best Local Similarity 40.0%; Pred. No. 1.9e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLDIV 9
Db 1 FLDIV 5

Search completed: May 1, 2003, 23:24:59
Job time : 29 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:24:07 ; Search time 46 Seconds
(without alignments)
16.882 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMPIQIV 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 32825 seqs, 8628665 residues

Total number of hits satisfying chosen parameters: 30581

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1 | 47.7 | 8 | 10 | US-09-935-682-35 | Sequence 35, Appl |
| 2 | 47.7 | 9 | 9 | US-09-277-074-25 | Sequence 25, Appl |
| 3 | 47.7 | 9 | 9 | US-10-116-846B-35 | Sequence 35, Appl |
| 4 | 45.5 | 5 | 9 | US-10-000-273-9 | Sequence 9, Appl |
| 5 | 45.5 | 6 | 9 | US-09-814-604-45 | Sequence 45, Appl |
| 6 | 45.5 | 6 | 9 | US-09-922-226-31 | Sequence 31, Appl |
| 7 | 45.5 | 6 | 9 | US-09-922-226-184 | Sequence 184, App |
| 8 | 45.5 | 6 | 9 | US-09-983-802-660 | Sequence 660, App |
| 9 | 45.5 | 9 | 10 | US-09-753-831-14 | Sequence 14, Appl |
| 10 | 43.2 | 8 | 9 | US-09-974-879-581 | Sequence 581, App |
| 11 | 43.2 | 9 | 9 | US-09-809-638-34 | Sequence 34, Appl |
| 12 | 43.2 | 9 | 9 | US-09-809-638-52 | Sequence 52, Appl |
| 13 | 43.2 | 9 | 9 | US-09-809-638-135 | Sequence 135, App |
| 14 | 43.2 | 9 | 9 | US-09-809-638-242 | Sequence 242, App |
| 15 | 43.2 | 9 | 9 | US-09-803-638-452 | Sequence 452, App |
| 16 | 43.2 | 9 | 9 | US-09-809-638-661 | Sequence 661, App |
| 17 | 43.2 | 9 | 9 | US-10-116-846B-31 | Sequence 31, Appl |
| 18 | 43.2 | 9 | 10 | US-09-780-053-33 | Sequence 33, Appl |
| 19 | 40.9 | 8 | 9 | US-10-033-741-17 | Sequence 17, Appl |

| | | | | | | |
|----|----|------|---|----|-------------------|-------------------|
| 20 | 18 | 40.9 | 8 | 9 | US-10-140-164-10 | Sequence 10, Appl |
| 21 | 18 | 40.9 | 8 | 9 | US-10-140-164-40 | Sequence 40, Appl |
| 22 | 18 | 40.9 | 9 | 1 | US-08-821-739A-48 | Sequence 48, Appl |
| 23 | 18 | 40.9 | 9 | 9 | US-09-826-290-24 | Sequence 24, Appl |
| 24 | 18 | 40.9 | 9 | 9 | US-09-277-074-35 | Sequence 35, Appl |
| 25 | 18 | 40.9 | 9 | 9 | US-09-791-389-98 | Sequence 98, Appl |
| 26 | 18 | 40.9 | 9 | 9 | US-09-791-389-98 | Sequence 98, Appl |
| 27 | 18 | 40.9 | 9 | 10 | US-09-897-107-21 | Sequence 21, Appl |
| 28 | 17 | 38.6 | 7 | 9 | US-09-895-793-969 | Sequence 969, App |
| 29 | 17 | 38.6 | 7 | 10 | US-09-822-827-969 | Sequence 969, App |
| 30 | 17 | 38.6 | 8 | 10 | US-09-756-283A-41 | Sequence 41, Appl |
| 31 | 17 | 38.6 | 9 | 1 | US-08-821-739A-50 | Sequence 50, Appl |
| 32 | 17 | 38.6 | 9 | 9 | US-10-017-327-3 | Sequence 3, Appl |
| 33 | 17 | 38.6 | 9 | 9 | US-10-017-327-5 | Sequence 5, Appl |
| 34 | 17 | 38.6 | 9 | 9 | US-10-094-699-79 | Sequence 79, Appl |
| 35 | 17 | 38.6 | 9 | 9 | US-10-094-699-80 | Sequence 80, Appl |
| 36 | 16 | 36.4 | 6 | 9 | US-09-727-963A-38 | Sequence 38, Appl |
| 37 | 16 | 36.4 | 6 | 10 | US-09-244-694-190 | Sequence 190, App |
| 38 | 16 | 36.4 | 7 | 9 | US-10-124-880-39 | Sequence 39, Appl |
| 39 | 16 | 36.4 | 8 | 9 | US-09-827-345-15 | Sequence 15, Appl |
| 40 | 16 | 36.4 | 8 | 9 | US-10-062-710-137 | Sequence 137, App |
| 41 | 16 | 36.4 | 8 | 10 | US-09-243-079-8 | Sequence 8, Appl |
| 42 | 16 | 36.4 | 8 | 10 | US-09-243-079-9 | Sequence 9, Appl |
| 43 | 16 | 36.4 | 8 | 10 | US-09-243-079-10 | Sequence 10, Appl |
| 44 | 16 | 36.4 | 8 | 10 | US-09-243-079-11 | Sequence 11, Appl |
| 45 | 16 | 36.4 | 8 | 10 | US-09-012-135A-18 | Sequence 18, Appl |

ALIGNMENTS

RESULT 1
US-09-935-682-35
Sequence 35, Application US/09935682
Patent No. US20020059032A1
GENERAL INFORMATION:
APPLICANT: Societe de Conseils de Recherches et d'Applications Scientifiques
APPLICANT: Ferrer, Camara Y.
TITLE OF INVENTION: Rational Selection of Putative Peptides from Identified Nucleot.
FILE REFERENCE: 58767.000005
CURRENT APPLICATION NUMBER: US/09/935,682
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 09/257,525
PRIOR FILING DATE: 1999-02-25
PRIOR APPLICATION NUMBER: PCT/FR00/00460
PRIOR FILING DATE: 2000-02-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35
LENGTH: 8
TYPE: PRT
ORGANISM: Homo sapiens
US-09-935-682-35
Query Match 47.7%; Score 21; DB 10; Length 8;
Best Local Similarity 80.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMP 5
Db 1 ISEMP 5
RESULT 2
US-09-277-074-25
Sequence 25, Application US/09277074
GENERAL INFORMATION:
APPLICANT: Sheehan, Linda A.
TITLE OF INVENTION: IN VIVO ACTIVATION OF TUMOR-SPECIFIC CYTOTOXIC T CELLS
FILE REFERENCE: SCR21555
CURRENT APPLICATION NUMBER: US/09/277,074

;; CURRENT FILING DATE: 1999-03-26
;; PRIOR APPLICATION NUMBER: 08/355,558
;; PRIOR FILING DATE: 1994-12-14
;; PRIOR APPLICATION NUMBER: PCT/US95/16415
;; PRIOR FILING DATE: 1995-12-14
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 25
;; LENGTH: 9
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-277-074-25

Query Match 47.7%; Score 21; DB 9; Length 9;
Best Local Similarity 37.5%; Pred. No. 2.9e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISMFLOI 8
Db 2 LNMFCOL 9

RESULT 3
US-10-116-846B-35
; Sequence 35, Application US/10116846B
; Publication No. US20030073102A1
; GENERAL INFORMATION:
; APPLICANT: KWOK, William
; APPLICANT: NEPOM, Gerald
; APPLICANT: GEBE, John
; APPLICANT: REIDONEN, Helena
; APPLICANT: LIU, Andrew
; TITLE OF INVENTION: METHODS OF MHC CLASS II EPIOTOPE MAPPING, DETECTION OF AUTOIMMUNE
; TITLE OF INVENTION: AND ANTIGENS, AND AUTOIMMUNE TREATMENT
; FILE REFERENCE: 20149-0005200S
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 60/282,328
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/308,962
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Herpes Simplex Virus 2
US-10-116-846B-35

Query Match 47.7%; Score 21; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 MFLQIY 9
Db 1 LFLMHLV 6

RESULT 4
US-10-000-273-9
; Sequence 9, Application US/10000273
; Patent No. US20020160386A1
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows 95
;; SOFTWARE: FastSeq for Windows Version 2.0b
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/000,273
;; FILING DATE: 02-No. US20020160386A1-2001
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/964,127
;; FILING DATE: 06-NOV-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Crews, Ph.D., L. Lee
;; REGISTRATION NUMBER: P-43,567
;; REFERENCE/DOCKET NUMBER: 07334/038001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-000-273-9

Query Match 45.5%; Score 20; DB 9; Length 5;
Best Local Similarity 80.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISMFL 5
Db 1 ISMFL 5

RESULT 5
US-09-814-604-45
; Sequence 45, Application US/09814604
; Publication No. US20030003517A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Elliott S.
; APPLICANT: Chandraratna Roshantha A.
; TITLE OF INVENTION: Methods of Detecting Dissociated Nuclear
; FILE REFERENCE: P-AR 4528
; CURRENT APPLICATION NUMBER: US/09/814,604
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-604-45

Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFLQIY 9
Db 1 LFLMHLV 6

RESULT 6
US-09-922-226-31
; Sequence 31, Application US/09922226
; Publication No. US2003007764A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi

```

; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening For Compounds That
; FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-31
Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Gaps 0;

QY 4 MFLQIY 9
;|:|:|:
DB 1 LFLVLF 6

RESULT 7
US-09-922-226-184
; Sequence 184, Application US/09922226
; Publication No. US2003007664A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi
; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening For Compounds That
; FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-184
Query Match 45.5%; Score 20; DB 9; Length 6;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 2; Conservative 4; Mismatches 0; Gaps 0;

QY 4 MFLQIY 9
;|:|:|:
DB 1 LFLVLF 6

RESULT 8
US-09-983-802-660
; Sequence 660, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 660
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-983-802-660

Query Match 45.5%; Score 20; DB 9; Length 9;
Best Local Similarity 57.1%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 7
DB 1 IREIFLR 7

RESULT 9

US-09-753-831-14
Sequence 14, Application US/09753831
Patent No. US20020137683A1
GENERAL INFORMATION:
APPLICANT: Hogan, Kevin T.
TITLE OF INVENTION: C-SKI Oncogene-Derived Peptides for Prevention,
FILE REFERENCE: 26747-27
CURRENT APPLICATION NUMBER: US/09/753,831
PRIOR FILING DATE: 2001-01-03
PRIOR APPLICATION NUMBER: U.S. 60/174296
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: No. US20020137683A1a peptide
US-09-753-831-14

Query Match 45.5%; Score 20; DB 10; Length 9;
Best Local Similarity 33.3%; Pred. No. 2.9e+05;
Matches 3; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
DB 1 LERSRVRY 9

RESULT 10

US-09-974-879-581
Sequence 581, Application US/09974879
Publication No. US20030028003A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 125 Human Secreted Proteins
FILE REFERENCE: P2020P2
CURRENT APPLICATION NUMBER: US/09/974,879
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/239,893
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 09/818,683
PRIOR FILING DATE: 2001-03-28
PRIOR APPLICATION NUMBER: US 09/305,736
PRIOR FILING DATE: 1999-05-05
PRIOR APPLICATION NUMBER: PCT/US98/23435
PRIOR FILING DATE: 1998-11-04
PRIOR APPLICATION NUMBER: US 60/064,911
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,912
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,983
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,900
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,988
PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: US 60/064,987
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,908
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,984
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,985
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/066,094
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,100
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,089
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,095
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: US 60/066,090
NUMBER OF SEQ ID NOS: 611
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 581
LENGTH: 8
TYPE: PRT
ORGANISM: Homo sapiens
US-09-974-879-581

Query Match 43.2%; Score 19; DB 9; Length 8;
Best Local Similarity 42.9%; Pred. No. 2.9e+05;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMFLQIY 9
DB 2 EIRGVF 8

RESULT 11

US-09-809-638-34
Sequence 34, Application US/09809638
Publication No. US2003005895A1
GENERAL INFORMATION:
APPLICANT: Mary Faris
APPLICANT: Pia M. Challita-Eid
APPLICANT: Steve Chappell Mitchell
APPLICANT: Daniel E.H. Afar
APPLICANT: Arthur B. Raitano
APPLICANT: Ava Jakobovits
TITLE OF INVENTION: 125PSC8: A TISSUE SPECIFIC PROTEIN
TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
FILE REFERENCE: 129.35US01
CURRENT APPLICATION NUMBER: US/09/809,638
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 746
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-809-638-34

Query Match 43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 25.0%; Pred. No. 2.9e+05;
Matches 2; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
DB 2 SEKYWCLF 9

RESULT 12
US-09-809-638-52
Sequence 52, Application US/09809638
Publication No. US2003005895A1
GENERAL INFORMATION:

```

; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-52

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
   :|:|:|
Db 2 LSPFLFTI 9

```

```

RESULT 13
US-09-809-638-135
; Sequence 135, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-135

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
   :|:|:|
Db 2 LSPFLFTI 9

```

```

RESULT 14
US-09-809-638-242
; Sequence 242, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN

```

```

; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-242

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 50.0%; Pred. No. 2.9e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ISEMFLQI 8
   :|:|:|
Db 2 LSPFLFTI 9

```

```

RESULT 15
US-09-809-638-452
; Sequence 452, Application US/09809638
; Publication No. US20030059895A1
; GENERAL INFORMATION:
; APPLICANT: Mary Paris
; APPLICANT: Pia M. Challita-Bid
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 125P5C8: A TISSUE SPECIFIC PROTEIN
; TITLE OF INVENTION: HIGHLY EXPRESSED IN VARIOUS CANCERS
; FILE REFERENCE: 129.35US01
; CURRENT APPLICATION NUMBER: US/09/809,638
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 746
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 452
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-638-452

```

```

Query Match          43.2%; Score 19; DB 9; Length 9;
Best Local Similarity 25.0%; Pred. No. 2.9e+05;
Matches 2; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 SEMFLQIY 9
   ||:|:|:|
Db 2 SEKYWKLP 9

```

```

Search completed: May 1, 2003, 23:32:01
Job time : 47 secs

```


GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:10:31 ; Search time 4.96429 Seconds
(without alignments)
1168,869 Million cell updates/sec

Title: US-09-658-621B-26
Perfect score: 44
Sequence: 1 ISEWFLQIV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 44 | 100.0 | 9 | 20 | US-09-658-621-26 |
| 2 | 44 | 100.0 | 10 | 20 | US-09-658-621-29 |
| 3 | 44 | 100.0 | 292 | 26 | US-10-221-279-12252 |
| 4 | 44 | 100.0 | 321 | 1 | PCT-US00-05882-861 |
| 5 | 44 | 100.0 | 321 | 23 | US-09-925-301-861 |
| 6 | 44 | 100.0 | 475 | 1 | PCT-US01-30151-3 |

| | | | | | | |
|----|----|-------|------|----|-----------------------|----------------------|
| 7 | 44 | 100.0 | 475 | 20 | US-09-658-621-67 | Sequence 67, Appl |
| 8 | 44 | 100.0 | 495 | 20 | US-09-658-621-2 | Sequence 2, Appl1 |
| 9 | 44 | 100.0 | 515 | 1 | PCT-US02-07826-212 | Sequence 212, App |
| 10 | 44 | 100.0 | 515 | 24 | US-10-097-340-212 | Sequence 212, App |
| 11 | 44 | 100.0 | 515 | 25 | US-10-171-311-156 | Sequence 156, App |
| 12 | 44 | 100.0 | 1255 | 1 | PCT-US02-19669-311 | Sequence 311, Appl |
| 13 | 44 | 100.0 | 1255 | 23 | US-09-996-069-10 | Sequence 10, Appl |
| 14 | 44 | 100.0 | 1255 | 25 | US-10-171-311-156 | Sequence 156, App |
| 15 | 44 | 100.0 | 1255 | 25 | US-10-171-311-156 | Sequence 311, Appl |
| 16 | 41 | 93.2 | 508 | 17 | US-09-366-670-20 | Sequence 20, Appl |
| 17 | 41 | 93.2 | 508 | 24 | US-10-057-136-20 | Sequence 20, Appl |
| 18 | 36 | 81.8 | 448 | 21 | US-09-760-446A-1839 | Sequence 1839, App |
| 19 | 36 | 81.8 | 1083 | 26 | US-10-206-664-1839 | Sequence 1839, App |
| 20 | 36 | 81.8 | 1083 | 21 | US-09-791-537-1101725 | Sequence 101725, App |
| 21 | 36 | 81.8 | 1083 | 21 | US-09-791-537-110555 | Sequence 110555, App |
| 22 | 36 | 81.8 | 1083 | 21 | US-09-791-537-110590 | Sequence 110590, App |
| 23 | 36 | 81.8 | 1091 | 21 | US-09-791-537-29746 | Sequence 29746, App |
| 24 | 36 | 81.8 | 1554 | 27 | US-60-389-987-1395 | Sequence 1395, App |
| 25 | 36 | 81.8 | 1554 | 27 | US-60-412-418-1395 | Sequence 1395, App |
| 26 | 34 | 77.3 | 545 | 23 | US-09-966-728-2 | Sequence 2, Appl1 |
| 27 | 33 | 75.0 | 57 | 19 | US-09-513-996A-78504 | Sequence 78504, A |
| 28 | 33 | 75.0 | 61 | 19 | US-09-513-996A-78166 | Sequence 78166, A |
| 29 | 33 | 75.0 | 71 | 19 | US-09-513-996A-14787 | Sequence 14787, A |
| 30 | 33 | 75.0 | 71 | 20 | US-09-620-394B-2999 | Sequence 2999, App |
| 31 | 33 | 75.0 | 71 | 23 | US-09-935-625-24194 | Sequence 24194, A |
| 32 | 33 | 75.0 | 381 | 21 | US-09-791-537-132877 | Sequence 132877, App |
| 33 | 33 | 75.0 | 381 | 21 | US-60-360-039-6373 | Sequence 6373, App |
| 34 | 33 | 75.0 | 839 | 21 | US-09-708-427-16579 | Sequence 16579, A |
| 35 | 33 | 75.0 | 898 | 21 | US-09-708-427-16579 | Sequence 16579, A |
| 36 | 33 | 75.0 | 955 | 21 | US-09-708-427-16577 | Sequence 16577, A |
| 37 | 33 | 75.0 | 1005 | 19 | US-09-513-996A-60874 | Sequence 60874, A |
| 38 | 33 | 75.0 | 1028 | 23 | US-09-935-625-9787 | Sequence 9787, App |
| 39 | 33 | 75.0 | 1057 | 19 | US-09-513-996A-60873 | Sequence 60873, A |
| 40 | 33 | 75.0 | 1080 | 23 | US-09-935-625-9786 | Sequence 9786, App |
| 41 | 33 | 75.0 | 1114 | 19 | US-09-513-996A-60872 | Sequence 60872, A |
| 42 | 33 | 75.0 | 1137 | 23 | US-09-935-625-9785 | Sequence 9785, App |
| 43 | 32 | 72.7 | 97 | 21 | US-09-758-441-244 | Sequence 244, App |
| 44 | 32 | 72.7 | 97 | 26 | US-10-211-725-244 | Sequence 244, App |
| 45 | 32 | 72.7 | 108 | 27 | US-60-128-476-2574 | Sequence 2574, App |

ALIGNMENTS

RESULT 1
US-09-658-621-26
Sequence 26, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offtinga, Rieken
APPLICANT: Wellet, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MGC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-26
Query Match 100.0%; Score 44; DB 20; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | |
Db 1 ISEMFLQIY 9

RESULT 2

US-09-658-621-29
; Sequence 29, Application US/09658621
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Heukamp, Lukas Carl
; APPLICANT: Offringa, Rienk
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/09/658,621
; CURRENT FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-658-621-29

Query Match 100.0%; Score 44; DB 20; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.093;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | |
Db 1 ISEMFLQIY 9

RESULT 3
US-10-221-279-12252
; Sequence 12252, Application US/10221279
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-046
; CURRENT APPLICATION NUMBER: US/10/221,279
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 12360
; SOFTWARE: Custom
; SEQ ID NO 12252
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-221-279-12252

Query Match 100.0%; Score 44; DB 26; Length 292;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | |
Db 110 ISEMFLQIY 118

RESULT 4

PCT-US00-05882-861
; Sequence 861, Application PC/TUS0005882
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Human Cancer Associated Gene Sequences and Polypeptides
; FILE REFERENCE: P4106PCT
; CURRENT APPLICATION NUMBER: PCT/US00/05882
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 861
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05882-861

Query Match 100.0%; Score 44; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | |
Db 139 ISEMFLQIY 147

RESULT 5
US-09-925-301-861
; Sequence 861, Application US/09925301
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P4106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 861
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-861

Query Match 100.0%; Score 44; DB 23; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | |
Db 139 ISEMFLQIY 147

RESULT 6
PCT-US01-30151-3
; Sequence 3, Application PC/TUS0130151
; GENERAL INFORMATION:

APPLICANT: Genesance Pharmaceuticals, Inc.
APPLICANT: Anne, Chew
APPLICANT: Koshi, Beena
TITLE OF INVENTION: HAPLOYPES OF THE MUC1 GENE
FILE REFERENCE: MUC1 WWH-1408PCT
CURRENT APPLICATION NUMBER: PCT/US01/30151
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/236,113
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-30151-3

Query Match 100.0%; Score 44; DB 1; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 293 ISEMPLOY 301

RESULT 7
US-09-658-621-67
Sequence 67, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 67
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-67

Query Match 100.0%; Score 44; DB 20; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 293 ISEMPLOY 301

RESULT 8
US-09-658-621-2
Sequence 2, Application US/09658621
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille

TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 495
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621-2

Query Match 100.0%; Score 44; DB 20; Length 495;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 313 ISEMPLOY 321

RESULT 9
PCT-US02-07826-212
Sequence 212, Application PC/TUS0207826
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030PC
CURRENT APPLICATION NUMBER: PCT/US02/07826
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 212
LENGTH: 515
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-07826-212

Query Match 100.0%; Score 44; DB 1; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9
Db 333 ISEMPLOY 341

RESULT 10
US-10-097-340-212
Sequence 212, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN

```

; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Michael E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VELIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-097-340-212

Query Match          100.0%; Score 44; DB 24; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ISEMFLQIY 9
        |||||
        333 ISEMFLQIY 341

RESULT 11
US-10-171-311-156
; Sequence 156, Application US/1017311
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersht, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
```

```

; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-156

Query Match          100.0%; Score 44; DB 25; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ISEMFLQIY 9
        |||||
        333 ISEMFLQIY 341

RESULT 12
PCT-US02-19669-311
; Sequence 311, Application PCT/US0219669
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc. et al.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038PC
; CURRENT APPLICATION NUMBER: PCT/US02/19669
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-19669-311

Query Match          100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ISEMFLQIY 9
        |||||
        1073 ISEMFLQIY 1081

RESULT 13
US-09-996-069-10
; Sequence 10, Application US/09996069
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBI
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: MO1015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; PRIOR FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 1255
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

Query Match
Best Local Similarity 100.0%; Score 44; DB 23; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081

RESULT 14
US-10-171-311-158
; Sequence 158, Application US/10171311
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatc, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Query Match
Best Local Similarity 100.0%; Score 44; DB 25; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081

RESULT 15
US-10-177-293-311
; Sequence 311, Application US/10177293
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatc, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos

```

```

; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-311

Query Match
Best Local Similarity 100.0%; Score 44; DB 25; Length 1255;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081

```

Search completed: May 1, 2003, 23:19:10
 Job time : 5.96429 secs

FT NON TER 5 5 5
SQ SEQUENCE 5 AA; 654 MW; 72CB19C9C0300000 CRC64;
Query Match 27.3%; Score 12; DB 1; Length 5;
Best Local Similarity 20.0%; Pred. No. 1.1e+05;
Matches 1; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 4 MFLQI 8
DB 1 IFFEV 5
RESULT 7
CIP2_MYTED STANDARD; PRT; 6 AA.
AC P13737;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Contraction-inhibiting peptide II (MIP II).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE.
RC TISSUE=Pedal ganglion;
RX MEDLINE=88240357; PubMed=3377776;
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.,
"Structures and actions of Mytilus inhibitory peptides.",
RL Biochem. Biophys. Res. Commun. 152:1376-1382(1998).
CC -1- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
MUSCLES.
KW -1- SIMILARITY: TO MIP I.
DR PIR; B27696; B27696.
KM Hormone; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA; 621 MW; 72C9C6876DD81000 CRC64;
Query Match 27.3%; Score 12; DB 1; Length 6;
Best Local Similarity 66.7%; Pred. No. 1.1e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 4 MFL 6
DB 4 MEV 6
RESULT 8
EIO1_LITRU STANDARD; PRT; 6 AA.
ID EIO1_LITRU
AC P82096;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Electrin 1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.,
"Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella.",
RL Aust. J. Chem. 52:639-645(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED.
KW Amphibian skin; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA; 792 MW; 6683704772C9A000 CRC64;

Query Match 27.3%; Score 12; DB 1; Length 6;
Best Local Similarity 40.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 5 FLQIY 9
DB 1 FVPIW 5

RESULT 9
ALU6_CYDPO STANDARD; PRT; 8 AA.
ID ALU6_CYDPO
AC P82157;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cydiastratin 6.
OS Cydia pomonella (Codling moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptysa; Tortricidae; Tortricidae; Olethreutinae; Cydia.
OX NCBI_TaxID=82600;
RN [1]
RP SEQUENCE.
RC TISSUE=Larva;
RX MEDLINE=98054519; PubMed=9392829;
RA Dve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
"Leptodipteran peptides of the allatostatin superfamily.",
RL Peptides 18:1301-1309(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 936 MW; 0B2879C458573767 CRC64;
Query Match 27.3%; Score 12; DB 1; Length 8;
Best Local Similarity 50.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 6 LQIY 9
DB 1 LPIY 4
RESULT 10
FAR2_CALVO STANDARD; PRT; 9 AA.
ID FAR2_CALVO
AC P41857;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Calliphrasmamide 2.
OS Calliphora vomitoria (Blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Calliphoridae; Calliphora.
OX NCBI_TaxID=27454;
RN [1]
RP SEQUENCE.
RC TISSUE=Thoracic ganglion;
RX MEDLINE=92196111; PubMed=1549595;
RA Dve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
Rahfeld J.F., Thorpe A.,
"Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT neuropeptides (designated calliphrasmamide) from the blowfly
RT Calliphora vomitoria.",
RL Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC -1- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC SALIVARY GLAND OF CALLIPHORA.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
DR PIR; B41978; B41978.

```
KM Neuropeptide; Amidation.
FT MOD RES 9 AA; 1128 MW; 29D00699CAB6C5A7 CRC64;
SQ SEQUENCE 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Query Match
Best Local Similarity 33.3%; Pred. No. 1.1e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 SEMFLO 7
|:|:|:
DB 3 SQDFMR 8

RESULT 11
FAR3 CALVO STANDARD; PRT; 9 AA.
AC P41658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE CalliFMRamide 3.
OS Calliphora vomitoria (Blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Calliphoridae; Calliphora.
OX NCBI_TaxID=27454;
RN [1]
RP SEQUENCE.
RC TISSUE=Thoracic ganglion;
RX MEDLINE=9219611; PubMed=1549595;
RA Duvé H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA Rehfeld J.F., Thorpe A.;
RT "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT neuropeptides (designated calliFMRamides) from the blowfly
RT Calliphora vomitoria."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC -1- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC SALIVARY GLAND OF CALLIPHORA.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRPAMIDE RELATED PEPTIDE)
CC FAMILY.
CC PIR; C41978; C41978.
KM Neuropeptide; Amidation.
FT MOD RES 9 AA; 1114 MW; 2F0B0699CAB6C5A7 CRC64;
SQ SEQUENCE 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Query Match
Best Local Similarity 33.3%; Pred. No. 1.1e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 SEMFLO 7
|:|:|:
DB 3 SQDFMR 8

RESULT 12
FLRF HIRME STANDARD; PRT; 4 AA.
AC P42561;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FLRFamide.
OS Hirudo medicinalis (Medicinal leech), and
OS Helisoma trivolvis (Snail).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
OC Arynchobdellida; Hirudinoformes; Hirudindae; Hirudo.
OX NCBI_TaxID=6421, 27815;
RN [1]
RP SEQUENCE.
RC SPECIES=H. medicinalis;
RX MEDLINE=9215953; PubMed=1686933;
RA Evans G.D., Pohl J., Kartsolis M.A., Calabrese R.L.;
RT "Identification of Rfamide neuropeptides in the medicinal leech."
```

```
RL Peptides 12:897-908(1991).
RN [2]
RP SEQUENCE.
RC SPECIES=H.trivolvis; TISSUE=Kidney;
RX MEDLINE=9428647; PubMed=7912428;
RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
RT "FMRamide-related peptides from the kidney of the snail, Helisoma
RT trivolvis."
RL Peptides 15:31-36(1994).
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRPAMIDE RELATED PEPTIDE)
CC FAMILY.
KM Neuropeptide; Amidation.
FT MOD RES 4 AA; 582 MW; 69D40729A0000000 CRC64;
SQ SEQUENCE 4 AA; 549 MW; 64540729A0000000 CRC64;

Query Match
Best Local Similarity 66.7%; Pred. No. 1.1e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 FLO 7
|:|:|:
DB 1 FLR 3

RESULT 13
FLRN ANTFL STANDARD; PRT; 4 AA.
AC PS8707;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Antho-RNamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nyantheae; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RP SEQUENCE, AND MASS-SPECTROMETRY.
RX MEDLINE=90319122; PubMed=1973541;
RA Grimmelikhuijzen C.J.P., Rinehart K.L. Jr., Jacob E., Graff D.,
RA Reincheid R.K., Norhacker H.-P., Staley A.L.;
RT "Isolation of U-3-phenylacetyl-Leu-Arg-Asn-NH2 (Antho-RNamide), a sea
RT anemone neuropeptide containing an unusual amino-terminal blocking
RT group."
RL Proc. Natl. Acad. Sci. U.S.A. 87:5410-5414(1990).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Neuron-specific.
CC -1- MASS SPECTROMETRY: MW=549.3; METHOD=FAB.
KM Neuropeptide; Amidation.
FT MOD RES 1 AA; 549 MW; 64540729A0000000 CRC64;
FT MOD RES 4 AA; 549 MW; 64540729A0000000 CRC64;
SQ SEQUENCE 4 AA; 549 MW; 64540729A0000000 CRC64;

Query Match
Best Local Similarity 25.0%; Score 11; DB 1; Length 4;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 FLO 7
|:|:|:
DB 1 FLR 3

RESULT 14
PSK DAUCA STANDARD; PRT; 5 AA.
AC PS8261;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phytosulfokine-alpha (PSK-alpha) [Contains: Phytosulfokine-beta (PSK-
DE beta)].
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```


CURRENT APPLICATION NUMBER: US/09/658,621B
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 29
LENGTH: 10
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-29

Query Match 100.0%; Score 44; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 1 ISEMFLQIY 9

RESULT 3
US-60-452-680-12836
Sequence 12836, Application US/60452680
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: GRUPE, Andrew
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001450
CURRENT APPLICATION NUMBER: US/60/452,680
CURRENT FILING DATE: 2003-03-07
NUMBER OF SEQ ID NOS: 116213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12836
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-452-680-12836

Query Match 100.0%; Score 44; DB 7; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 4
US-60-453-135-8037
Sequence 8037, Application US/60453135
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: IAKOBOVA, Olga
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001456
CURRENT APPLICATION NUMBER: US/60/453,135
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8037
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-135-8037

Query Match 100.0%; Score 44; DB 7; Length 264;

Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 5
US-60-453-050-8037
Sequence 8037, Application US/60453050
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: LUKE, May
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001457
CURRENT APPLICATION NUMBER: US/60/453,050
CURRENT FILING DATE: 2003-03-10
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8037
LENGTH: 264
TYPE: PRT
ORGANISM: Homo sapiens
US-60-453-050-8037

Query Match 100.0%; Score 44; DB 7; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 82 ISEMFLQIY 90

RESULT 6
US-09-658-621B-67
Sequence 67, Application US/09658621B
GENERAL INFORMATION:
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Heukamp, Lukas Carl
APPLICANT: Offringa, Rienk
APPLICANT: Melief, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
TITLE OF INVENTION: WOC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621B
CURRENT FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 67
LENGTH: 475
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-67

Query Match 100.0%; Score 44; DB 5; Length 475;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
| | | | | | | | | |
| | | | | | | | | |
Db 293 ISEMFLQIY 301

```
RESULT 7
US-10-348-119-309
; Sequence 309, Application US/10348119
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 NP
; CURRENT APPLICATION NUMBER: US/10/348,119
; CURRENT FILING DATE: 2003-01-17
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 309
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-348-119-309

Query Match
Best Local Similarity 100.0%; Score 44; DB 6; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 8
US-60-422-176-59
; Sequence 59, Application US/60422176
; GENERAL INFORMATION:
; APPLICANT: Pharmacia Corporation
; TITLE OF INVENTION: DIFFERENTIALLY EXPRESSED GENES INVOLVED IN CANCER, THE POLYPEPTIDE
; TITLE OF INVENTION: ENCODED THEREBY, AND METHODS OF USING THE SAME
; FILE REFERENCE: 01040 PR
; CURRENT APPLICATION NUMBER: US/60/422,176
; CURRENT FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 3114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 475
; TYPE: PRT
; ORGANISM: homo sapiens
US-60-422-176-59

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 9
US-60-452-680-12835
; Sequence 12835, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12835
; LENGTH: 475
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-12835

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 10
US-60-453-135-8036
; Sequence 8036, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8036
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-8036

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 11
US-60-453-050-8036
; Sequence 8036, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8036
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-8036

Query Match
Best Local Similarity 100.0%; Score 44; DB 7; Length 475;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
Db 293 ISEMFLQTY 301

RESULT 12
US-09-658-621b-2
; Sequence 2, Application US/09658621B
; GENERAL INFORMATION:
```

```
APPLICANT: Taylor-Papadimitriou, Joyce
APPLICANT: Haukamp, Lukas Carl
APPLICANT: Offringa, Rijk
APPLICANT: Mellet, Cornelis Johanna Maria
APPLICANT: Acres, Bruce
APPLICANT: Thomas, Mireille
TITLE OF INVENTION: MUC-1 derived peptides
FILE REFERENCE: 029395-017
CURRENT APPLICATION NUMBER: US/09/658,621B
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: US 60/187,215
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: GB 9921242.5
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: EP 99 40 2237.4
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 495
TYPE: PRT
ORGANISM: Homo sapiens
US-09-658-621B-2
```

```
Query Match          100.0%; Score 44; DB 5; Length 495;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 313 ISEMFLQIY 321
```

```
RESULT 13
PCT-US02-19669A-311
Sequence 311, Application PC/TUS0219669A
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
FILE REFERENCE: MRI-038PC
CURRENT APPLICATION NUMBER: PCT/US02/19669A
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/239,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 311
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-19669A-311
```

```
Query Match          100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081
```

```
RESULT 14
```

```
PCT-IL02-00255-2
Sequence 2, Application PC/TL0200255
GENERAL INFORMATION:
APPLICANT: Ramot University Authority for Applied Research
APPLICANT: WRESCHNER, Daniel
APPLICANT: YOSHI-LERNER, Meirav
APPLICANT: SMORODINSKY, Nechama
TITLE OF INVENTION: Peptides and Antibodies to MUC 1 Proteins
FILE REFERENCE: P-3891-PC
CURRENT APPLICATION NUMBER: PCT/IL02/00255
PRIOR FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: 60/279,408
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1255
TYPE: PRT
ORGANISM: Homo sapiens
PCT-IL02-00255-2
```

```
Query Match          100.0%; Score 44; DB 1; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081
```

```
RESULT 15
US-09-949-016-9618
Sequence 9618, Application US/09949016
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CD001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9618
LENGTH: 1041
TYPE: PRT
ORGANISM: Human
US-09-949-016-9618
```

```
Query Match          81.8%; Score 36; DB 5; Length 1041;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ISEMFLQIY 9
Db 679 ISEMFLQIY 687
```

```
Search completed: May 1, 2003, 23:21:01
Job time : 1.91071 secs
```

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:14:31 ; Search time 28 Seconds
(without alignments)
66.229 Million cell updates/sec

Title: US-09-658-621b-26

Perfect score: 44

Sequence: 1 ISEWFLQIV 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 686

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_plant:*

10: sp_todent:*

11: sp_virus:*

12: sp_vertebrate:*

13: sp_unclassified:*

14: sp_virus:*

15: sp_bacteriaph:*

16: sp_archaeap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 18 | 40.9 | 9 | 2 | P72345 |
| 2 | 16 | 36.4 | 8 | 2 | O9RQ49 |
| 3 | 15 | 34.1 | 8 | 8 | O94V76 |
| 4 | 15 | 34.1 | 9 | 7 | O78226 |
| 5 | 15 | 34.1 | 9 | 8 | O94X56 |
| 6 | 15 | 34.1 | 9 | 8 | O94X56 |
| 7 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 8 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 9 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 10 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 11 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 12 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 13 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 14 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 15 | 15 | 34.1 | 9 | 8 | O94NB1 |
| 16 | 15 | 34.1 | 9 | 8 | O94NB1 |

| | | | | | | |
|----|----|------|---|----|--------|---------------------|
| 17 | 13 | 29.5 | 9 | 4 | P78484 | P78484 homo sapien |
| 18 | 13 | 29.5 | 9 | 4 | O14277 | O14277 homo sapien |
| 19 | 13 | 29.5 | 9 | 11 | O62530 | O62530 mus spreus |
| 20 | 13 | 29.5 | 9 | 12 | O88953 | O88953 vaccinia vi |
| 21 | 12 | 27.3 | 8 | 2 | O45889 | O45889 clostridium |
| 22 | 12 | 27.3 | 8 | 2 | O938P2 | O938P2 pseudomonas |
| 23 | 12 | 27.3 | 8 | 2 | O37854 | O37854 bacterioph |
| 24 | 12 | 27.3 | 8 | 10 | O40530 | O40530 nicotiana t |
| 25 | 12 | 27.3 | 9 | 2 | O47556 | O47556 escherichia |
| 26 | 12 | 27.3 | 9 | 2 | P83157 | P83157 anabaena sp |
| 27 | 12 | 27.3 | 9 | 4 | O9UKJ6 | O9UKJ6 homo sapien |
| 28 | 12 | 27.3 | 9 | 4 | O9UC36 | O9UC36 homo sapien |
| 29 | 12 | 27.3 | 9 | 11 | O9QWC2 | O9QWC2 mus musculi |
| 30 | 12 | 27.3 | 9 | 12 | O83622 | O83622 mus musculi |
| 31 | 12 | 27.3 | 9 | 12 | O83622 | O83622 mus musculi |
| 32 | 11 | 25.0 | 7 | 2 | O54248 | O54248 streptomyces |
| 33 | 11 | 25.0 | 7 | 5 | P83274 | P83274 macrobrachi |
| 34 | 11 | 25.0 | 8 | 2 | O9RQ57 | O9RQ57 buchnera ap |
| 35 | 11 | 25.0 | 8 | 2 | O56429 | O56429 thermus the |
| 36 | 11 | 25.0 | 8 | 2 | P72279 | P72279 rhodococcus |
| 37 | 11 | 25.0 | 8 | 4 | O15890 | O15890 homo sapien |
| 38 | 11 | 25.0 | 8 | 4 | O16428 | O16428 homo sapien |
| 39 | 11 | 25.0 | 8 | 5 | O9UB13 | O9UB13 albinaria h |
| 40 | 11 | 25.0 | 8 | 5 | P83275 | P83275 macrobrachi |
| 41 | 11 | 25.0 | 8 | 5 | P83316 | P83316 penaeus mon |
| 42 | 11 | 25.0 | 8 | 6 | O9MYL5 | O9MYL5 pongo pygma |
| 43 | 11 | 25.0 | 8 | 8 | O9XNP8 | O9XNP8 boophilus m |
| 44 | 11 | 25.0 | 8 | 8 | O9GD47 | O9GD47 hydraestele |
| 45 | 11 | 25.0 | 8 | 8 | O9GC24 | O9GC24 nenga puni |

ALIGNMENTS

| | | | | |
|-----------------------|---|---|-------|---------------|
| RESULT 1 | P72345 | PRELIMINARY; | PRT; | 9 AA. |
| ID | P72345 | 01-FEB-1997 (TREMBLrel. 02, Created) | | |
| AC | P72345 | 01-JAN-1998 (TREMBLrel. 05, Last sequence update) | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last annotation update) | | | |
| DE | HYPOTHETICAL 1.0 kDa protein (Fragment). | | | |
| GN | TABA. | | | |
| OS | Pseudomonas syringae. | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; | | | |
| OX | Pseudomonas. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=93167809; PubMed=7679566; | | | |
| RA | Barta T.M., Kinschert T.G., Uchytel T.F., Willis D.K.; | | | |
| RT | "DNA sequence and transcriptional analysis of the tda gene required for tabtoxin biosynthesis by Pseudomonas syringae." | | | |
| RL | Appl. Environ. Microbiol. 59:458-466(1993). | | | |
| DR | EMBL; S54909; AAB25381.2; - | | | |
| KW | HYPOTHETICAL protein. | | | |
| FT | NON TER | | | |
| SQ | SEQUENCE 9 AA; 1037 MW; 2B34D9D5B805B047 CRC64; | | | |
| Query Match | 40.9%; | Score 18; | DB 2; | Length 9; |
| Best Local Similarity | 80.0%; | Pred. No. 6.7e+05; | | |
| Matches | 4; | Conservative | 0; | Mismatches 1; |
| Indels | 0; | Gaps | 0; | |
| DB | 3 ISEWFLQIV | | | |
| QY | 1 ISEWFLQIV | | | |
| DB | 3 ISEWFLQIV | | | |
| RESULT 2 | O9RQ49 | PRELIMINARY; | PRT; | 8 AA. |
| ID | O9RQ49 | 01-MAY-2000 (TREMBLrel. 13, Created) | | |
| AC | O9RQ49 | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Created) | | | |

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE NIFS protein homolog (Fragment).
 GN NIFS.
 OS Buchnera aphidicola.
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 CX NCBI_TaxId=9;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20022990; PubMed=10555290;
 RA Clark M.A., Moran N.A., Baumann P.;
 RT "Sequence evolution in bacterial endosymbionts having extreme base
 RT compositions";
 RL Mol. Biol. Evol. 16:1586-1598(1999).
 DR EMBL; AF130814; AAF13805.1; -.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 992 MW; F3A73B5047672336 CRC64;

Query Match 36.4%; Score 16; DB 2; Length 8;
 Best Local Similarity 66.7%; Pred. No. 6.7e+05;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 MFLQIY 9
 DB 1 MKLPIY 6

RESULT 3
 Q94VF6 PRELIMINARY; PRT; 8 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Cytochrome c oxidase subunit I (Fragment).
 GN COI.
 OS Varanus jobiensis.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Anguimorpha; Varanidae; Varanus.
 CX NCBI_TaxId=169843;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Ast J.C.;
 RT "Mitochondrial DNA evidence and evolution in Varanoidea (Squamata).";
 RL Cladistics 17:0-0(2001).
 DR EMBL; AF407507; AAL10075.1; -.
 KW Mitochondrion.
 FT NON TER 8
 SQ SEQUENCE 8 AA; 1144 MW; EFD729DB436411A6 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 8;
 Best Local Similarity 100.0%; Pred. No. 6.7e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 FLQ 7
 DB 6 FLQ 8

RESULT 4
 Q78226 PRELIMINARY; PRT; 9 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Lymphocyte antigen (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN=CRO437; TISSUE=KIDNEY;
 RX MEDLINE=88084418; PubMed=3692165;
 RA Golubic M., Budimir O., Schoepfer R., Kasahara M., Mayer W.E.,
 RA Figueroa F., Klein J.;
 RT "Nucleotide sequence analysis of class II genes borne by mouse t
 RT chromosomes";
 RL Genet. Res. 50:137-146(1987).
 DR EMBL; J38590; AAB57294.1; -.
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1176 MW; 86CB1412C729C33A CRC64;

Query Match 34.1%; Score 15; DB 7; Length 9;
 Best Local Similarity 60.0%; Pred. No. 6.7e+05;
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFL 6
 DB 1 SDRFL 5

RESULT 5
 Q94XE6 PRELIMINARY; PRT; 9 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Cytochrome c oxidase subunit III (Fragment).
 GN COX3.
 OS Tectocoris diophthalmus (cotton harlequin bug).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Paraneoptera; Hemiptera; Eumecoptera;
 OC Pentatomidae; Pentatomomorpha; Pentatomidae;
 OC Tectocoridae; Tectocoris.
 CX NCBI_TaxId=159956;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21396409; PubMed=11504862;
 RT "Increased rate of gene rearrangement in the mitochondrial genomes of
 RT three orders of hemipteroid insects";
 RL Mol. Biol. Evol. 18:1828-1832(2001).
 DR EMBL; AF335990; AAK55283.1; -.
 KW Mitochondrion.
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1206 MW; A2C563636B5041A6 CRC64;

Query Match 34.1%; Score 15; DB 8; Length 9;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOIY 9
 DB 1 YMTIY 5

RESULT 6
 Q94NB2 PRELIMINARY; PRT; 9 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Cytochrome oxidase subunit III (Fragment).
 GN COIII.
 OS Microtus murinus (lesser mouse lemur).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
 OC Microcebus.

QY 1 ISEMFLQIY 9
:|||||
Db 729 ISEMFLQIY 737

RESULT 5

D95283
Probable Arac-type regulator [imported] - Sinorhizobium meliloti (strain 1021) megaplasm
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95283
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bove
; Kaimari, S.; Keating, D.H.; Salin, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A>Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11461432
A:Accession: D95283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <RUR>
A:Cross-references: GB:AE006469; PIDN:AAK64830.1; PID:G14523242; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kaimari, S.; Keating, D.H.; Kles, E.; Komp, C.; LaLaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Genetics: annotation
A:Cross-references:
A:Gene: SMO0319
A:Genome: plasmid

Query Match 75.0%; Score 33; DB 2; Length 333;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
:|||||
Db 65 SDMFQVY 72

RESULT 6

T29826
Hypothetical protein C5SH.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29826
R:Pauley, A.; Gattung, S.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans comid C5SH.
A:Reference number: Z20694
A:Accession: T29826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-381 <PAU>
A:Cross-references: EMBL:U55567; PIDN:AAA97988.1; GSPDB:GN00023; CESP:C5SH.2
A:Experimental source: strain Bristol N2; clone C5SH1
A:Genetics:
A:Gene: CESP:C5SH.2
A:Map position: 5
A:Intons: 21/2; 57/2; 103/3; 156/2; 182/2; 238/3; 268/3; 320/1; 347/2
C:Superfamily: GTP-binding regulatory protein Gs alpha chain

Query Match 75.0%; Score 33; DB 2; Length 381;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
:|||||
Db 169 MSEMFLKIF 177

RESULT 7
A47447
Calcium channel protein alpha-1 chain (variant doe-1) - electric ray (Discopyge ommata)

C:Species: Discopyge ommata
C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A47447
R:Horne, W.A.; Ellinger, P.T.; Imman, I.; Zhou, M.; Tsiem, R.W.; Schwarz, T.L.
Proc. Natl. Acad. Sci. U.S.A. 90, 3787-3791, 1993
A>Title: Molecular diversity of Ca(2+) channel alpha 1 subunits from the marine ray Disc
A:Reference number: A47447; MUID:93248175; PMID:7683405
A:Accession: A47447
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2223 <HOR>
A>Note: sequence extracted from NCBI backbone (NCBIP:130671)
C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 75.0%; Score 33; DB 2; Length 2223;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
:|||||
Db 535 MSEMFLKIV 542

RESULT 8

T13980
Calcium channel alpha-1 chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13980
R:Smith, L.A.; Wang, X.J.; Peixoto, A.P.; Neumann, E.K.; Hall, L.M.; Hall, J.C.
J. Neurosci. 16, 7868-7879, 1996
A>Title: A Drosophila calcium channel alpha-1 subunit gene maps to a genetic locus ass
A:Reference number: Z17844; MUID:97141514; PMID:8987815
A:Accession: T13980
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1851 <SMI>
A:Cross-references: EMBL:U55776; NID:G1737063; PID:G1737064; PIDN:AA47406.1
A:Genetics:
A:Gene: cala
A:Cross-references: FlyBase:FBgn0005563
C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 1851;
Best Local Similarity 55.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
:|||||
Db 485 MSEMFLKIV 493

RESULT 9

A41098
Calcium channel protein alpha-1 chain isoform A - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 24-Sep-1999
C:Accession: A41098; A35901
R:Starr, T.V.B.; Pryce, W.; Snutch, T.P.
Proc. Natl. Acad. Sci. U.S.A. 88, 5621-5625, 1991
A>Title: Primary structure of a calcium channel that is highly expressed in the rat cere
A:Reference number: A41098; MUID:91288516; PMID:1648226
A:Accession: A41098
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2212 <GBA>
A:Cross-references: GB:M64373; NID:G203110; PIDN:AAA40806.1; PID:G203111
R:Snutch, T.P.; Leonard, J.P.; Gilbert, M.M.; Lester, H.A.; Davidson, N.

```
Proc. Natl. Acad. Sci. U.S.A. 87, 3391-3395, 1990
A>Title: Rat brain expresses a heterogeneous family of calcium channels.
A/Reference number: A35901; MUID:90239020; PMID:1692134
A/Accession: A35901
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A/Molecule type: mRNA
A/Residues: 1435-1667 <SNV>
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2212;
Best Local Similarity 55.6%; Pred. No. 2.2e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ISEMFLQIY 9
Db 533 MSEMFLQIY 541

RESULT 10
146477
Calcium channel BI-1 - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: 146477; 146478
R/Motl, Y.; Friedrich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman
Nature 350, 398-402, 1991
A>Title: Primary structure and functional expression from complementary DNA of a brain c
A/Reference number: 146477; MUID:91187110; PMID:1849233
A/Accession: 146477
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2273 <MOR>
A/Cross-references: EMBL:X57476; NID:g1522; PIDD:CAA40714.1; PID:g1523
A/Accession: 146478
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1856, 'R', 1860-1862, 'GV', 1865-1866, 'VIS', 1870-1876, 'K', 1878-1879, '
A/Cross-references: EMBL:X57688; NID:g1524; PIDD:CAA40871.1; PID:g1525
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2273;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ISEMFLQIY 9
Db 531 MSEMFLQIY 539

RESULT 11
146480
Calcium channel BI-2 - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C/Accession: 146480; 146479
R/Motl, Y.; Friedrich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Bosse, E.; Hofman
Nature 350, 398-402, 1991
A>Title: Primary structure and functional expression from complementary DNA of a brain c
A/Reference number: 146477; MUID:91187110; PMID:1849233
A/Accession: 146480
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2242 <MOR>
A/Cross-references: EMBL:X57689; NID:g1528; PIDD:CAA40872.1; PID:g1529
A/Accession: 146479
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1856, 'L', 1858, 'R', 1860-1862, 'AM', 1865-1866, 'HMP', 1870-1876, 'N', 1878-1879, '
A/Cross-references: EMBL:X57477; NID:g1526; PIDD:CAA40715.1; PID:g1527
C/Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match 72.7%; Score 32; DB 2; Length 2424;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;
```

```
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ISEMFLQIY 9
Db 531 MSEMFLQIY 539

RESULT 12
157007
Pregnancy-specific glycoprotein - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1995 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C/Accession: 157007
R/Rudert, F.; Sanders, A.M.; Thompson, J.A.; Rebstock, S.; Zimmermann, W.A.
Mamm. Genome 3, 262-273, 1992
A>Title: Characterization of murine carcinoembryonic antigen gene family members.
A/Reference number: 157007; MUID:92345715; PMID:1638085
A/Accession: 157007
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-141 <RBS>
A/Cross-references: GB:M83341; NID:g200309; PIDD:AAA39914.1; PID:g554253
C/Genetics:
A/Gene: CEM7
A/Insertion: 22/1
C/Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-term
C/Keywords: glycoprotein
F/137/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEA1>
F/48-96/Domain: immunoglobulin homology <IMW>

Query Match 70.5%; Score 31; DB 2; Length 141;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 2 SEMFLQIY 9
Db 134 SEMFLQIY 141

RESULT 13
S23659
Superoxide dismutase (EC 1.15.1.1) (Mn) - European spiny lobster (fragment)
C/Species: Palinurus vulgaris (European spiny lobster)
C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C/Accession: S23659
R/Smith, M.W.; Doolittle, R.F.
J. Mol. Evol. 34, 175-184, 1992
A>Title: A comparison of evolutionary rates of the two major kinds of superoxide dismuta
A/Reference number: S23655; MUID:92211732; PMID:1556751
A/Accession: S23659
A/Molecule type: mRNA
A/Residues: 1-144 <SMI>
A/Cross-references: EMBL:X64063; NID:g10095; PIDD:CAA45419.1; PID:g1335662
C/Genetics:
A/Genome: nuclear
A/Function:
A/Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C/Superfamily: superoxide dismutase (Mn)
C/Keywords: manganese; metalloprotein; mitochondrion; oxidoreductase

Query Match 70.5%; Score 31; DB 2; Length 144;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 ISEMFLQIY 9
Db 2 ISEMFLQIY 10

RESULT 14
Q08BNS
Integral membrane protein - saimiriine herpesvirus 1 (strain 11)
C/Species: saimiriine herpesvirus 1
```


Job time : 25 secs

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
 OX NCBI_TaxID=4039;
 RN [1]
 RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.
 RC STRAIN=cv. US-Harumaki90sun;
 RX MEDLINE=20212743; PubMed=10750705;
 RA Hanai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,
 RA Kamada H., Sakagami Y.;
 RT "A secreted peptide growth factor, phytosulfokine, acting as a
 RT stimulatory factor of carrot somatic embryo formation.";
 RL Plant Cell Physiol. 41:27-32(2000).
 CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE
 CC CELLS. BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC
 CC EMBRYOS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- PTM: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A
 CC PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOKINE FAMILY.
 KM Growth factor; Sulfation. PHYTOSULFOKINE-BETA.
 FT PEPTIDE 1 4
 FT MOD_RES 1 1 SULFATION.
 FT MOD_RES 3 3 SULFATION.
 SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

 Query Match 25.0%; Score 11; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.1e+05; Indels 0; Gaps 0;
 Matches 2; Conservative 0; Mismatches 0;

 QY 8 IV 9
 Db 2 IV 3

 RESULT 15
 RE21 LITRU STANDARD; PRT; 5 AA.
 ID RE21 LITRU
 AC P82071;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Rubellidin 2.1.
 OS Litoria rubella (Desert tree frog).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
 OC Pelodyadinae; Litoria.
 OX NCBI_TaxID=104895;
 RN [1]
 RP SEQUENCE, AND MASS SPECTROMETRY.
 RC TISSUE=Skin secretion;
 RC Steinhorn S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
 RA Tyler M.J., Wallace J.C.;
 RT "The structure of new peptides from the Australian red tree frog
 RT 'Litoria rubella', the skin peptide profile as a probe for the study
 RT of evolutionary trends of amphibians.";
 RL Aust. J. Chem. 49:955-963(1996).
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTI-BIOTIC
 CC ACTIVITY.
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
 CC -1- MASS SPECTROMETRY: MW=626; METHOD=PAB.
 KM Amphibian skin.
 SQ SEQUENCE 5 AA; 626 MW; 6DD9C9CB10300000 CRC64;

 Query Match 25.0%; Score 11; DB 1; Length 5;
 Best Local Similarity 66.7%; Pred. No. 1.1e+05;
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 3 EMF 5
 Db 2 ERF 4

Search completed: May 1, 2003, 23:24:03

A:Accession: A35175
 A:Molecule type: mRNA
 A:Residues: 1-952,1033-1344 <LIG1>
 A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:
 A:Experimental source: splice form A
 A:Note: Genbank entries HUMESP1A1 and HUMESP1A2 present only the amino- and carboxyl-ter
 A:Accession: B35175
 A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <LIG2>
 A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
 A:Experimental source: splice form B
 A:Note: Genbank entries HUMESP1B1 and HUMESP1B2 present only the amino- and carboxyl-ter
 J:Gendler, S.J.; Lannas, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Pest, N.; Burchel
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel
 A:Reference number: A35886; MUID:90368715; PMID:1697369
 A:Accession: A35886
 A:Molecule type: mRNA
 A:Status: not compared with conceptual translation
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:G188869; PIDN:AAA59876.1; PID:g188870
 A:Note: Genbank entry HUMMUCAB includes one copy of the tandemly repeated sequence
 R:Ilan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716; PMID:2394722
 A:Accession: A35887
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
 A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:g189599
 A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated sequ
 R:Wreschner, D.H.; Hareven, M.; Tzarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413; PMID:2351132
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <MRB>
 A:Cross-references: EMBL:X52229; NID:g37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <MRZ
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
 R:Abu, M.; Siddiqui, J.; Kute, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473; PMID:2597151
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142,'O',144-162,'O',164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
 R:Masuoka, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglut
 A:Reference number: U00235; MUID:9313189; PMID:1478919
 A:Accession: U00235
 A:Molecule type: mRNA
 A:Residues: 998-1011,'E',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
 A:Experimental source: gastric carcinoma cell
 R:Zirhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 356, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
 A:Reference number: S51026; MUID:95080414; PMID:7989707
 A:Accession: S51026
 A:Contents: annotation
 A:Note: underlined tyrosine residues in the carboxyl-terminal non-repetitive region ar
 C:Comment: this protein is length polymorphic. Individuals may have between 21 and 125 c
 partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: SdkI and CheoI nuclease sites in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C:Genetics:
 A:Gene: GDB:MUC1; PUM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q23-1q23
 A:Intron: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 C:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphs
 F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRA>
 F:1-62/Region: mucin 1 amino-terminal non-repetitive
 F:1-1344/Region: signal sequence #link PRA #status predicted <SIGA>
 F:1-19,29-32/Domain: signal sequence #link PRA #status predicted <SIGB>
 F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PRB>
 F:1-19,29-212,1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PRB>
 F:138-1017/Region: 20-residue repeats (GSTAPPAGVTSAPDRPAP)
 F:143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1243-1272/Domain: transmembrane #status predicted <TM>
 F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted
 Query Match 100.0%; Score 44; DB 1; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMFLOQY 9
 Db 1162 ISEMFLOQY 1170
 RESULT 3
 S54293
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 01-Sep-1995
 C:Accession: S54293
 R:Homma, Y.; Emori, Y.
 EMBO J. 14, 286-291, 1995
 A:Title: A dual functional signal mediator showing rhoGAP and phospholipase C-delta str
 A:Reference number: S54293; MUID:95137008; PMID:7835339
 A:Accession: S54293
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1083 <ROM>
 A:Cross-references: EMBL:D31962
 Query Match 81.8%; Score 36; DB 2; Length 1083;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ISEMFLOQY 9
 Db 721 ISEMFLOQY 729
 RESULT 4
 G59435
 DUC-1 (deleted in liver cancer), p12 [imported] - huamn
 C:Species: Homo sapiens (man)
 C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
 C:Accession: G59435
 R:Yuan, B.Z.; Miller, M.J.; Keck, C.L.; Zimonjic, D.B.; Thorgeirsson, S.S.; Popescu, N
 Cancer Res. 58, 2196-2199, 1998
 A:Title: Cloning, characterization, and chromosomal localization of a gene frequently de
 A:Reference number: G59435
 A:Accession: G59435
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1091 <YUA>
 A:Cross-references: GB:NP_006085; PID:g5174427; PIDN:NP_006085.1
 Query Match 81.8%; Score 36; DB 2; Length 1091;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:08:21 ; Search time 0.946429 Seconds
(without alignments)
914.184 Million cell updates/sec

Title: US-09-658-621B-26
Perfect score: 44
Sequence: 1 ISEMFLQIT 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 segs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 44 | 100.0 | 255 | 1 S48146 | mucin 1 precursor, |
| 2 | 44 | 100.0 | 1344 | 1 A5175 | mucin 1 precursor, |
| 3 | 36 | 81.8 | 1083 | 2 S54293 | regulator protein |
| 4 | 36 | 81.8 | 1091 | 2 G59435 | DIC-1 (deleted in |
| 5 | 33 | 75.0 | 333 | 2 D55283 | probable Arac-type |
| 6 | 33 | 75.0 | 381 | 2 T29826 | hypothetical prote |
| 7 | 33 | 75.0 | 2223 | 2 A47447 | calcium channel pr |
| 8 | 32 | 72.7 | 1851 | 2 T13980 | calcium channel al |
| 9 | 32 | 72.7 | 2212 | 2 A41098 | calcium channel pr |
| 10 | 32 | 72.7 | 2273 | 2 T46477 | calcium channel BI |
| 11 | 32 | 72.7 | 2424 | 2 T46480 | calcium channel BI |
| 12 | 31 | 70.5 | 1441 | 2 T57007 | calcium channel BI |
| 13 | 31 | 70.5 | 144 | 2 T57007 | pregnancy-specific |
| 14 | 31 | 70.5 | 366 | 1 Q08BNS | superoxide dismuta |
| 15 | 31 | 70.5 | 426 | 2 T41682 | integral membrane |
| 16 | 31 | 70.5 | 477 | 2 T04916 | hypothetical prote |
| 17 | 31 | 70.5 | 491 | 2 C70868 | hypothetical prote |
| 18 | 31 | 70.5 | 492 | 2 T23502 | hypothetical prote |
| 19 | 31 | 70.5 | 606 | 2 H64986 | hypothetical prote |
| 20 | 31 | 70.5 | 896 | 2 S76064 | hypothetical prote |
| 21 | 31 | 70.5 | 896 | 2 S59980 | phycoerythrin anch |
| 22 | 31 | 70.5 | 974 | 2 T04910 | hypothetical prote |
| 23 | 31 | 70.5 | 995 | 2 H59432 | Ribop protein hom |
| 24 | 31 | 70.5 | 1071 | 2 S38164 | ATP-binding protei |
| 25 | 31 | 70.5 | 1531 | 1 S420X1 | major merozoite su |
| 26 | 31 | 70.5 | 1639 | 2 S05603 | probable major sur |
| 27 | 31 | 70.5 | 1640 | 2 A24594 | GCN1 homolog - fis |
| 28 | 31 | 70.5 | 2570 | 2 T37919 | thyronine |
| 29 | 30 | 68.2 | 117 | 2 C71811 | thyronine |

| | | | | | |
|----|----|------|-----|----------|---------------------|
| 30 | 30 | 68.2 | 244 | 2 H69844 | diadenosine tetrap |
| 31 | 30 | 68.2 | 332 | 2 T41227 | hypothetical prote |
| 32 | 30 | 68.2 | 334 | 2 F69469 | conserved hypothet |
| 33 | 30 | 68.2 | 383 | 2 A13224 | two component sens |
| 34 | 30 | 68.2 | 385 | 2 T24328 | hypothetical prote |
| 35 | 30 | 68.2 | 390 | 2 D81289 | probable aminotran |
| 36 | 30 | 68.2 | 408 | 2 S76830 | hypothetical prote |
| 37 | 30 | 68.2 | 416 | 2 AF2070 | hypothetical prote |
| 38 | 30 | 68.2 | 579 | 2 AC2200 | potassium-transport |
| 39 | 30 | 68.2 | 630 | 2 A39344 | tumor-associated m |
| 40 | 30 | 68.2 | 631 | 2 T52257 | epistatin - mouse |
| 41 | 30 | 68.2 | 770 | 2 S37962 | mitochondrial inte |
| 42 | 30 | 68.2 | 815 | 2 S67675 | probable membrane |
| 43 | 30 | 68.2 | 946 | 2 A71843 | D-lactate dehydrog |
| 44 | 30 | 68.2 | 948 | 2 F64672 | D-lactate dehydrog |
| 45 | 30 | 68.2 | 981 | 2 S55132 | hypothetical prote |

ALIGNMENTS

RESULT 1

S48146
mucin 1 precursor, non-repetitive splice form Y [validated] - human
N/Alternate names: breast carcinoma-associated DF3 antigen; epistatin; MUC1/Y protein;
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C/Accession: S48146
R/Rizhan-Licht, S.; Vos, H.L.; Baruch, A.; Elroy-Stein, O.; Sagiv, D.; Keydar, I.; Hilk
Bur, J. Biochem. 224, 787-795, 1994
A/Title: Characterization and molecular cloning of a novel MUC1 protein, devoid of tand
A/Reference number: S48146; PMID:95010060; PMID:7925397
A/Accession: S48146
A/Molecule type: mRNA
A/Residues: 1-255 <ZRI>
A/Cross-reference: EMBL:X60761; NID:9541679; PIDN:CA56734.1; PID:9541680
A/Comment: For alternative splice forms with a tandem repeat domain, see PIR:A5175.

C/Genetics:
A/Gene: GDB:MUC1; PUM
A/Cross-reference: GDB:120705; OMIM:158340
A/Map position: 1q21-1q23
A/Intons: 20/1; 53/3; 73/3; 95/2; 141/1; 181/3; 231/3
C/Suprafamily: polymorphic epithelial mucin
C/Keywords: glycoprotein; phosphoprotein; polymorphism; transmembrane protein
F/1-53/Region: mucin 1 amino-terminal non-repetitive
F/1-23/Domain: signal sequence #status predicted <SIG>
F/24-255/Product: mucin 1, non-repetitive splice form Y #status predicted <MAT>
F/154-255/Region: mucin 1 carboxyl-terminal non-repetitive
F/156-183/Domain: transmembrane #status predicted <TRM>
F/155,133/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 44; DB 1; Length 255;
Best local similarity 100.0%; Pred. No. 0.05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIT 9
DB 73 ISEMFLQIT 81

RESULT 2

A5175
mucin 1 precursor, repetitive splice form A [validated] - human
N/Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistatin
ncretic mucin; polymorphic epithelial mucin (PEM)
N/Contents: mucin 1 precursor; epithelial tumor antigen splice form; mucin 1 precursor,
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C/Accession: A5175; B35175; A35887; S10572; S40293; A36735; EX0066; S10218; S51
R/Ligtenberg, M.J.L.; Vos, H.L.; Gemmessen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A/Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A/Reference number: A5175; PMID:9020794; PMID:2318825

A>Note: host Saimiri sciureus (common squirrel monkey)
 C>Date: 31-Dec-1992 #sequence_revision 51-Dec-1992 #text_change 16-Jul-1999
 C:Accession: A36810
 R:Albrecht, J.
 submitted to the EMBL Data Library, January 1992
 A:Description: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A36806
 A:Accession: A36810
 A:Molecule type: DNA
 A:Residues: 1-356 <ALB>
 A:Cross-references: GB:X6346; NID:G60320; PIDN:CAA45662.1; PID:G60360
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W
 J:Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A37309; WUID:9233688; PMID:1321287
 A:Contents: annotation; possible protein-coding frames
 A>Note: neither amino acid nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 39
 C:Superfamily: cytomegalovirus UL100 protein
 C:Keywords: transmembrane protein
 F:18-34/Domain: transmembrane #status predicted <TM1>
 F:87-103/Domain: transmembrane #status predicted <TM2>
 F:152-168/Domain: transmembrane #status predicted <TM3>
 F:214-230/Domain: transmembrane #status predicted <TM4>
 F:236-252/Domain: transmembrane #status predicted <TM5>
 F:274-290/Domain: transmembrane #status predicted <TM6>
 F:305-321/Domain: transmembrane #status predicted <TM7>

Query Match 70.5%; Score 31; DB 1; Length 366;
 Best Local Similarity 66.7%; Pred. No. 55;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOIY 9
 ::|||
 Db 259 MTEMFLODY 267

RESULT 15
 T41682
 Hypothetical protein SPCP1E11.03 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
 C:Accession: T41682
 R:Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21742
 A:Accession: T41682
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-426 <MED>
 A:Cross-references: EMBL:AL117483; PIDN:CA854862.1; GSPDB:GNO0068; SPDB:SPCP1E11.03
 A:Experimental source: strain 972h-; clone p1 pIR1
 C:Genetics:
 A:Gene: SPDB:SPCP1E11.03
 A:Map position: 3
 A:Introns: 217/2; 300/3; 411/2
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPCP1E11.03

Query Match 70.5%; Score 31; DB 2; Length 426;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ISEMPLOIY 9
 ::|||
 Db 127 IKEMFLDPY 135

Search completed: May 1, 2003, 23:13:53
 Job time : 2.94643 secs

Thu May 8 16:14:19 2003

us-09-658-621b-26.rspt

Page 6

Db 533 MSEMERKMY 541

Search completed: May 1, 2003, 23:12:50
Job time : 4.91071 secs

Query Match 75.0%; Score 33; DB 10; Length 955;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
Db 920 SQUELFQVY 927

RESULT 12

O9EN19 PRELIMINARY; PRT; 290 AA.
AC O9EN19;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AMV029.
GN Amasacta moorei entomopoxvirus (AmEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=28321;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20396580; PubMed=10936094;
RA Bawden A.L., Glasberg K.J., Diggins J., Shaw R., Farmerie W.,
Moyer R.W.;
RT "Complete Genomic Sequence of the Amasacta moorei Entomopoxvirus:
RT Analysis and Comparison with Other Poxviruses.";
RT Virology 274:120-139 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bawden A.L., Glasberg K.J., Diggins J., Shaw R., Farmerie W.,
Moyer R.W.;
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF250284; AAG02735.1; -;
DR InterPro; IPR001810; P-box.
DR PROSITE; PS50181; FBOX; 1.
SQ SEQUENCE 290 AA; 35839 MW; 709D59A7F0DB8C79 CRC64;

Query Match 72.7%; Score 32; DB 12; Length 290;
Best Local Similarity 55.6%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 207 LSEQFLFIY 215

RESULT 13

O9UG55 PRELIMINARY; PRT; 765 AA.
AC O9UG55;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE FORFI.
GN TT virus.
OC Viruses; ssDNA viruses, unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20417334; PubMed=10963344;
RA Tanaka Y., Orito E., Onno T., Nakano T., Hayashi K., Kato T.,
RA Mukaiide M., Iida S., Mizokami M.;
RT "Identification of a 23kDa protein encoded by putative open reading
RT frame 2 of TT virus (TTV) genotype 1 different from the other
RT genotypes.";
RL Arch. Virol. 145:1385-1398 (2000).
DR EMBL; AB030485; BAA90409.1; -;
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 765 AA; 90275 MW; 707029B845829B5F CRC64;

Query Match 72.7%; Score 32; DB 12; Length 765;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 731 TREMFLOTH 739

RESULT 14

O9USY1 PRELIMINARY; PRT; 2015 AA.
AC O9USY1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Microtubule-associated protein CP224.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=20233854; PubMed=10769206;
RA Graef R., Daubert C., Schliwa M.;
RT "Dictyostelium DdCP224 is a microtubule-associated protein and a
RT permanent centrosomal resident involved in centrosome duplication.";
RT J. Cell Sci. 113:1747-1758 (2000).
DR EMBL; AJ012088; CAB56504.1; -;
DR InterPro; IPR000357; HEAT repeat.
DR PROSITE; PS50077; HEAT_REPEAT; 1.
SQ SEQUENCE 2015 AA; 224041 MW; BA64B982ADC92BE CRC64;

Query Match 72.7%; Score 32; DB 5; Length 2015;
Best Local Similarity 62.5%; Pred. No. 3.8e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
Db 184 SELFLEIY 191

RESULT 15

O8RSW7 PRELIMINARY; PRT; 2327 AA.
AC O8RSW7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CAV2.1.
GN CACNA1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6N;
RX MEDLINE=21850703; PubMed=11756409;
RA Tsunemi T., Saegusa H., Ishikawa K., Nagayama S., Murakoshi T.,
RA Mizusawa H., Tanabe T.;
RT "Novel Cav2.1 Splice Variants Isolated from Purkinje Cells Do Not
RT Generate F-type Ca²⁺ Current.";
RT J. Biol. Chem. 277:7214-7221 (2002).
DR EMBL; AB066608; BAB85611.1; -;
SQ SEQUENCE 2327 AA; 263346 MW; 21FMA3297E7893C2 CRC64;

Query Match 72.7%; Score 32; DB 11; Length 2327;
Best Local Similarity 55.6%; Pred. No. 4.4e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:07:31 ; Search time 1.91071 Seconds
(without alignments)
970,540 Million cell updates/sec

Title: US-09-658-621b-26

Perfect score: 44

Sequence: 1 ISEMFLOIY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MNC:*
8: SP_ORGANELLE:*
9: SP_PLANT:*
10: SP_PLANT:*
11: SP_PODENT:*
12: SP_VIRUS:*
13: SP_VIRTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 41 | 93.2 | 553 | 6 Q9MZL1 | Q9MZL1 macaca mla |
| 2 | 36 | 81.8 | 1083 | 11 Q63744 | Q63744 rattus ratc |
| 3 | 36 | 81.8 | 1084 | 11 Q6RS41 | Q6RS41 mus musculu |
| 4 | 36 | 81.8 | 1091 | 4 Q43199 | Q43199 homo sapien |
| 5 | 36 | 81.8 | 1091 | 4 Q96GB1 | Q96GB1 homo sapien |
| 6 | 36 | 81.8 | 1554 | 4 Q9CGB0 | Q9CGB0 homo sapien |
| 7 | 33 | 75.0 | 182 | 10 Q8RX11 | Q8RX11 arabidopsis |
| 8 | 33 | 75.0 | 193 | 11 Q6Q408 | Q6Q408 cavia cutie |
| 9 | 33 | 75.0 | 330 | 5 Q9UJG3 | Q9UJG3 caenorhabd |
| 10 | 33 | 75.0 | 333 | 16 Q93OM3 | Q93OM3 rhizobium m |
| 11 | 33 | 75.0 | 955 | 10 Q48585 | Q48585 arabidopsis |
| 12 | 32 | 72.7 | 290 | 12 Q9EN19 | Q9EN19 amasacta moo |
| 13 | 32 | 72.7 | 765 | 12 Q9JGS5 | Q9JGS5 tt virus. p |
| 14 | 32 | 72.7 | 2015 | 5 Q9U5Y1 | Q9U5Y1 dictyostel |
| 15 | 32 | 72.7 | 2327 | 11 Q8RSW7 | Q8RSW7 mus musculu |
| 16 | 32 | 72.7 | 2365 | 11 Q8RSW6 | Q8RSW6 mus musculu |

| | | | | | |
|----|----|------|------|-----------|--------------------|
| 17 | 32 | 72.7 | 2472 | 4 Q9NS89 | Q9NS89 homo sapien |
| 18 | 32 | 72.7 | 2506 | 4 Q9NS88 | Q9NS88 homo sapien |
| 19 | 31 | 70.5 | 119 | 11 Q08928 | Q08928 mus musculu |
| 20 | 31 | 70.5 | 277 | 16 Q8RA94 | Q8RA94 thermaner |
| 21 | 31 | 70.5 | 426 | 3 Q9U84 | Q9U84 schizosacch |
| 22 | 31 | 70.5 | 437 | 5 Q8PPW6 | Q8PPW6 parametium |
| 23 | 31 | 70.5 | 450 | 17 Q8TME0 | Q8TME0 mechanosarc |
| 24 | 31 | 70.5 | 455 | 11 P70161 | P70161 mus musculu |
| 25 | 31 | 70.5 | 477 | 10 Q49640 | Q49640 arabidopsis |
| 26 | 31 | 70.5 | 492 | 5 Q93867 | Q93867 caenorhabd |
| 27 | 31 | 70.5 | 510 | 10 Q8W024 | Q8W024 arabidopsis |
| 28 | 31 | 70.5 | 599 | 10 Q8W4D7 | Q8W4D7 arabidopsis |
| 29 | 31 | 70.5 | 605 | 10 Q8SFU8 | Q8SFU8 arabidopsis |
| 30 | 31 | 70.5 | 612 | 17 Q8TMD2 | Q8TMD2 mechanosarc |
| 31 | 31 | 70.5 | 697 | 5 Q9N632 | Q9N632 leishmania |
| 32 | 31 | 70.5 | 896 | 16 Q55544 | Q55544 synchocyst |
| 33 | 31 | 70.5 | 974 | 10 Q49634 | Q49634 arabidopsis |
| 34 | 31 | 70.5 | 995 | 4 Q9Y3M8 | Q9Y3M8 homo sapien |
| 35 | 31 | 70.5 | 1056 | 11 Q923Q2 | Q923Q2 mus musculu |
| 36 | 31 | 70.5 | 1326 | 5 Q9W1Q5 | Q9W1Q5 drosophila |
| 37 | 30 | 68.2 | 117 | 16 Q92DB0 | Q92DB0 helicobacte |
| 38 | 30 | 68.2 | 141 | 11 Q9D740 | Q9D740 mus musculu |
| 39 | 30 | 68.2 | 203 | 5 Q76189 | Q76189 trypanosoma |
| 40 | 30 | 68.2 | 235 | 12 Q82936 | Q82936 jerry sloug |
| 41 | 30 | 68.2 | 235 | 12 Q98669 | Q98669 san angelo |
| 42 | 30 | 68.2 | 235 | 12 Q89796 | Q89796 jamestown c |
| 43 | 30 | 68.2 | 244 | 10 Q31614 | Q31614 bacillus su |
| 44 | 30 | 68.2 | 277 | 2 Q55036 | Q55036 synchocyst |
| 45 | 30 | 68.2 | 332 | 3 Q94472 | Q94472 schizosacch |

ALIGNMENTS

| RESULT 1 | Q9MZL1 | PRELIMINARY; | PRT; | 553 AA. |
|-----------------------|--|--------------|------|---------|
| ID | Q9MZL1 | | | |
| AC | Q9MZL1 | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Created) | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Last sequence update) | | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | | |
| DE | Mucin 1 (Fragment) | | | |
| OS | Macaca mulatta (Rhesus macaque) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; | | | |
| OC | Cercopithecoidea; Macaca. | | | |
| OX | NCBI_Taxid=9544; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=20330533; PubMed=1086975; | | | |
| RA | Vaughan H.A., Ho D.W.M., Karanikas V., Sandrin M.S., McKenzie I.F.C., | | | |
| RT | Pietersz G.A., | | | |
| RT | "The Immune Response of Mice and Cynomolgus Monkeys to Macaque Mucin1- | | | |
| RT | Mannan." | | | |
| RL | Vaccine 18:3297-3309(2000). | | | |
| DR | EMBL; AF176947; AAF82403.1; - | | | |
| DR | InterPro; IPR001064; Crystalin. | | | |
| DR | InterPro; IPR000082; SEA_domain. | | | |
| DR | Pfam; PF01390; SEA; 1. | | | |
| DR | SMART; SMO0200; SEA; 1. | | | |
| DR | PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1. | | | |
| DR | PROSITE; PS50024; SEA; 1. | | | |
| FT | NON_TER | | | |
| FT | NON_TER | | | |
| FT | NON_TER | | | |
| SQ | SEQUENCE 553 AA; 55778 MW; 60786DD2EE929318 CRC64; | | | |
| Query Match | 93.2%; Score 41; DB 6; Length 553; | | | |
| Best Local Similarity | 88.9%; Pred. No. 1.3; | | | |
| Matches | 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0; | | | |
| Oy | 1 ISEMFLOIY 9 | | | |
| Db | 396 ISEMFLOIY 404 | | | |

RESULT 2

ID 063744 PRELIMINARY; PRT; 1083 AA.

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE RhogAP.
OS Rattus rattus (Black rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10117;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=95137008; PubMed=7835339;

RT "A dual functional signal mediator showing RhogAP and phospholipase C-
delta stimulating activities";
RL EMBL J.14:286-291(1995).
DR EMBL; D31962; BAA21675.1; -.
DR HSSP; Q07960; IRGP.

DR InterPro; IPR001005; Myb_DNA_binding.

DR InterPro; IPR00198; RhGAP.

DR InterPro; IPR001660; SAM.

DR InterPro; IPR002913; START.

DR Pfam; PF00620; RhogAP.1.

DR Pfam; PF01852; START.1.

DR SMART; SM00324; RhogAP.1.

DR SMART; SM00454; SAM.1.

DR SMART; SM00234; START.1.

DR PROSITE; PS00037; MYB.1; UNKNOWN.1.

SQ SEQUENCE 1083 AA; 122465 MW; 076DDE77ACD9D012 CRC64;

SQ

Query Match 81.8%; Score 36; DB 11; Length 1083;

Best Local Similarity 77.8%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9

Db 721 ISEMPLOY 729

RESULT 3

ID 08R541 PRELIMINARY; PRT; 1084 AA.

AC 08R541;

DT 01-JUN-2002 (TREMblrel. 21, Created)

DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE D1C-1

GN ARHGAP7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RX MEDLINE=20169180; PubMed=10702663;

RYuan B.Z., Yang Y., Keck-Maggoner C.L., Zimonjic D.B.,

RA "Assignment and cloning of mouse Arhgap7 to chromosome 8A4-B2, a

RT conserved syntenic region of human chromosome 9p22--p21.";

RL Cytogenet. Cell Genet. 87:189-190(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RA Durlin M.E., Yuan B.Z., Thorgerisson S.S., Popescu N.C.;

RT "Gene structure, tissue expression and linkage mapping of the mouse

D1C-1 gene (Arhgap7).";

RL Gene 0:0-0(2002).

DR EMBL; AF411442; AAL87620.1; -.
DR EMBL; AF411435; AAL87620.1; JOINED.
DR EMBL; AF411436; AAL87620.1; JOINED.
DR EMBL; AF411437; AAL87620.1; JOINED.
DR EMBL; AF411438; AAL87620.1; JOINED.
DR EMBL; AF411439; AAL87620.1; JOINED.
DR EMBL; AF411440; AAL87620.1; JOINED.
DR EMBL; AF411441; AAL87620.1; JOINED.
SQ SEQUENCE 1084 AA; 122418 MW; 9FB1E4176ACABDEB CRC64;

SQ

Query Match 81.8%; Score 36; DB 11; Length 1084;

Best Local Similarity 77.8%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9

Db 722 ISEMPLOY 730

RESULT 4

ID 043199 PRELIMINARY; PRT; 1091 AA.

AC 043199;

DT 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)

DE Deleted in liver cancer-1.

DE D1C-1

GN Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Yuan B.Z., Miller M.J., Keck C.L., Zimonjic D.B., Thorgerisson S.S.,

RYuan B.Z., Miller M.J., Keck C.L., Zimonjic D.B., Thorgerisson S.S.,

RA Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF035119; AAB87700.1; -.
DR HSSP; Q07960; IRGP.

DR InterPro; IPR001005; Myb_DNA_binding.

DR InterPro; IPR00198; RhGAP.

DR InterPro; IPR001660; SAM.

DR InterPro; IPR002913; START.

DR Pfam; PF00620; RhogAP.1.

DR Pfam; PF01852; START.1.

DR SMART; SM00324; RhogAP.1.

DR SMART; SM00454; SAM.1.

DR SMART; SM00234; START.1.

DR PROSITE; PS00037; MYB.1; UNKNOWN.1.

SQ SEQUENCE 1091 AA; 122816 MW; 51712DE7ECDD0F52A CRC64;

SQ

Query Match 81.8%; Score 36; DB 4; Length 1091;

Best Local Similarity 77.8%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMPLOY 9

Db 729 ISEMPLOY 737

RESULT 5

ID 0960B1 PRELIMINARY; PRT; 1091 AA.

AC 0960B1; 014868;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Deleted in liver cancer 1 (HP protein).

DE D1C1 OR HP.

GN Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jeong S.-J., Ditchchev A., Lerman M., Ditchchilo A., Jung M.;
RT "Identification of HP/Dlci exon and introns.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-1091 FROM N.A.
RC TISSUE=LUNG;
RA Wei M.-H., Pack S., Ivanov S., Lerman M.I.;
RT "Cloning and Molecular Characterization of the Human Ortholog of the
RT Rat Dual Regulator p122RhoGAP.";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF408781; AAK97501.1; -.
DR EMBL; AF408768; AAK97501.1; JOINED.
DR EMBL; AF408770; AAK97501.1; JOINED.
DR EMBL; AF408772; AAK97501.1; JOINED.
DR EMBL; AF408773; AAK97501.1; JOINED.
DR EMBL; AF408774; AAK97501.1; JOINED.
DR EMBL; AF408775; AAK97501.1; JOINED.
DR EMBL; AF408776; AAK97501.1; JOINED.
DR EMBL; AF408777; AAK97501.1; JOINED.
DR EMBL; AF408778; AAK97501.1; JOINED.
DR EMBL; AF408779; AAK97501.1; JOINED.
DR EMBL; AF408780; AAK97501.1; JOINED.
DR EMBL; AF026219; AAB81637.1; -.
DR HSSP; Q07960; IRGP.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR001198; RhoGAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhoGAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhoGAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1091 AA; 122858 MW; BF9CD0B20A80AE83 CRC64;

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 1091;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEMFLOIY 9
Db 729 LSEFLOIY 737

RESULT 6
OQCE0 PRELIMINARY; PRT; 1554 AA.
AC OQCE0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1723 protein (Fragment).
GN KIAA1723.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=1082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL; AB051510; BAB21814.1; -.
DR HSSP; Q07960; IRGP.

```

```

DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR001198; RhoGAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhoGAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhoGAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 1554 AA; 173549 MW; 76FD31F139F2E12 CRC64;

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 1554;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEMFLOIY 9
Db 1192 LSEFLOIY 1200

RESULT 7
OBRXL1 PRELIMINARY; PRT; 182 AA.
AC OBRXL1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 20.4 kDa protein (Fragment).
GN ATG644750.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCB1_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.W., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.; Full length cDNA Clones.
RT "Arabidopsis Full length cDNA Clones."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY080825; AAL87302.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 182 AA; 20423 MW; 05160C0CFC3B562 CRC64;

Query Match
Best Local Similarity 75.0%; Score 33; DB 10; Length 182;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 SEMFLOIY 9
Db 153 SOLFLOIY 160

RESULT 8
OQ0408 PRELIMINARY; PRT; 193 AA.
AC OQ0408;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Epithelial mucin (Fragment).
GN MUC1.
OS Cavia culteri (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.

```

```
OX NCBI_TaxID=10144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ACTIVATING MAMMARY GLAND;
RX MEDLINE=96351712; PubMed=8747930;
RT Spicer A.P., Dunig T., Chilton B.S., Gendler S.J.;
RT "Analysis of mammalian MUC1 genes reveals potential functionally
RT important domains.";
RL Mamm. Genome 6:885-888(1995).
DR EMBL; L41546; AAB48542.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
RT NON TER
SQ SEQUENCE 193 AA; 21661 MW; D2593E4BE9FEC12F CRC64;

Query Match 75.0%; Score 33; DB 11; Length 193;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
ID 09UJG3 PRELIMINARY; PRT; 330 AA.
AC 09UJG3.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE F40D4.2 protein.
F40D4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditodea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX Matthews L.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RL "Genome sequence of the nematode C. elegans: A platform for
RL investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81536; CAB63199.1; -.
DR InterPro; IPR003003; 7TM_chemo2.
DR InterPro; IPR000168; 7TM_nematode.
DR Pfam; PF01604; 7tm 5; 1_nematode.
SQ SEQUENCE 330 AA; 37476 MW; C9B131FF6887A7AF CRC64;

Query Match 75.0%; Score 33; DB 5; Length 330;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQI 8
ID 0930M3 PRELIMINARY; PRT; 333 AA.
AC 0930M3.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative Arac-type regulator.
```

```
GN RA0172 OR SMA0319.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid psyma (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Sinorhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fleher R.F., Jones T., Komp C., Abola A.P.,
RA Barlet-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti psyma megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AB007210; AAK64830.1; -.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF00165; HTH_Arac; 2.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_2.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 333 AA; 36935 MW; 8BB2D702BD58A500 CRC64;

Query Match 75.0%; Score 33; DB 16; Length 333;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
ID 65 SDMEVQVY 72
DB 65 SDMEVQVY 72

RESULT 11
O48585 PRELIMINARY; PRT; 955 AA.
ID O48585
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein (SIMILARITY to DNA-damage-inducible protein
DE p).
GN T19K24.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kervlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=9156233; PubMed=10048488;
RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RL "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RL Sequence features of the regions of 1,081,958 bp covered by seventeen
RT physically assigned P1 and TAC clones.";
RL DNA Res 5:379-391(1998).
DR EMBL; AC002342; AAC78145.1; -.
DR EMBL; AB016874; BAB08828.1; -.
DR InterPro; IPR001126; UNIC_1like.
DR Pfam; PF00817; IMS; 1.
SQ SEQUENCE 955 AA; 104984 MW; BF01A0B417AAA355 CRC64;
```

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:19:16 ; Search time 24 Seconds
(without alignments)
15.554 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEMFLQIV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: * 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 231

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 15 | 34.1 | 6 | 1 | CPD1_MYTED |
| 2 | 15 | 34.1 | 8 | 1 | CPD1_ENTFA |
| 3 | 14 | 31.8 | 8 | 1 | AL18_CARMA |
| 4 | 13 | 29.5 | 9 | 1 | ULAE_HUMAN |
| 5 | 12 | 27.3 | 5 | 1 | E104_LITRU |
| 6 | 12 | 27.3 | 5 | 1 | UC22_MALZE |
| 7 | 12 | 27.3 | 6 | 1 | CIP2_MYTED |
| 8 | 12 | 27.3 | 6 | 1 | E101_LITRU |
| 9 | 12 | 27.3 | 8 | 1 | AL16_CYPDPO |
| 10 | 12 | 27.3 | 9 | 1 | PAR2_CALVO |
| 11 | 12 | 27.3 | 9 | 1 | PAR3_CALVO |
| 12 | 11 | 25.0 | 4 | 1 | FLNF_HIRME |
| 13 | 11 | 25.0 | 4 | 1 | FLRN_ANTEI |
| 14 | 11 | 25.0 | 5 | 1 | PSK_DAVCA |
| 15 | 11 | 25.0 | 5 | 1 | RE21_LITRU |
| 16 | 11 | 25.0 | 5 | 1 | RE31_LITRU |
| 17 | 11 | 25.0 | 7 | 1 | E105_LITRU |
| 18 | 11 | 25.0 | 7 | 1 | PAR1_HELTI |
| 19 | 11 | 25.0 | 7 | 1 | PAR1_PROCL |
| 20 | 11 | 25.0 | 7 | 1 | PAR2_PROCL |
| 21 | 11 | 25.0 | 7 | 1 | HY_PIG |
| 22 | 11 | 25.0 | 8 | 1 | ANG2_BOTJA |
| 23 | 11 | 25.0 | 8 | 1 | PAR1_PANRE |
| 24 | 11 | 25.0 | 8 | 1 | PAR3_HOMAM |
| 25 | 11 | 25.0 | 8 | 1 | PAR4_HOMAM |
| 26 | 11 | 25.0 | 8 | 1 | UPAA_HOMAM |
| 27 | 11 | 25.0 | 8 | 1 | PAR2_PANRE |
| 28 | 11 | 25.0 | 9 | 1 | PAR2_PANRE |
| 29 | 11 | 25.0 | 9 | 1 | FLN2_TREH |
| 30 | 11 | 25.0 | 9 | 1 | PGAR_DIAAB |
| 31 | 11 | 25.0 | 9 | 1 | ULAD_HUMAN |
| 32 | 10 | 22.7 | 4 | 1 | FYRI_ANTEI |
| 33 | 10 | 22.7 | 7 | 1 | LANC_CARUI |

| | | | | | | | |
|----|----|------|---|---|------------|--------|--------------|
| 34 | 10 | 22.7 | 7 | 1 | MMW1_ACHFU | P35919 | achalina fu |
| 35 | 10 | 22.7 | 7 | 1 | MMW3_ACHFU | P35921 | achalina fu |
| 36 | 10 | 22.7 | 8 | 1 | AL12_CARMA | P81815 | carcinus ma |
| 37 | 10 | 22.7 | 8 | 1 | NPB_BOVIN | P15507 | bos taurus |
| 38 | 10 | 22.7 | 9 | 1 | WOSF_CLYDA | P19853 | clypeaster |
| 39 | 10 | 22.7 | 9 | 1 | NEOU_CAVPO | P34966 | cavia porce |
| 40 | 10 | 22.7 | 9 | 1 | NEOU_HUMAN | P04277 | homo sapien |
| 41 | 10 | 22.7 | 9 | 1 | OXYT_BUPRE | P42995 | buto. regula |
| 42 | 10 | 22.7 | 9 | 1 | OXYT_CYPCA | P23879 | cypinus ca |
| 43 | 10 | 22.7 | 9 | 1 | OXYT_RABIT | P32878 | oryctolagus |
| 44 | 10 | 22.7 | 9 | 1 | OXYV_SQUAC | P43000 | squalus aca |
| 45 | 10 | 22.7 | 9 | 1 | UNI9_CLOPA | P81355 | clostridium |

ALIGNMENTS

RESULT 1
ID CIP1_MYTED STANDARD; PRT; 6 AA.
AC P13736;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Contraction-inhibiting peptide I (MIP I).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxId=6550;
RN [1]
RP SEQUENCE.
RC TISSUE=pedal ganglion;
RX MEDLINE=88240357; PubMed=3377776;
RA Hiraata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.,
RT "Structures and actions of Mytilus inhibitory peptides.";
RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC MUSCLES.
CC -!- SIMILARITY: TO MIP II.
DR PIR; A27696; A27696.
KW Hormone; Amidation.
FT MOD RES 6
SQ SEQUENCE 6 AA; 637 MW; 72C9C6875B81000 CRC64;
SO
Query Match 34.1%; Score 15; DB 1; Length 6;
Best Local Similarity 60.0%; Pred. No. 1.1e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 SEMFL 6
DB 2 SPMFV 6
RESULT 2
ID CIP1_ENTFA STANDARD; PRT; 8 AA.
AC P13269;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Sex pheromone CPD1.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxId=1351;
RN [1]
RP SEQUENCE.
RX MEDLINE=85040388; PubMed=6436978;
RA Suzuki A., Mori M., Sakakami Y., Isogai A., Fujino M., Kitada C.,
RT Craig R.A., Clewell D.B.;
RT "Isolation and structure of bacterial sex pheromone, cpd1.";
RT Science 226:849-850(1984).
CC -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC BACTERIOPHAGE PLASMID PPD1.

KM Pheromone. 8 AA; 913 MW; 8665B729C662C729 CRC64;
SQ SEQUENCE

Query Match 34.1%; Score 15; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 MFL 6
DB 4 MFL 6

RESULT 3

AL18_CARMA STANDARD; PRT; 8 AA.
ID AL18_CARMA
AC P81821;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Carcinus maenas.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Brachyura; Eubrachyura; Portunoidae; Portunidae; Carcinus.
OK NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RA MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas";
RL Ent. J. Biochem. 250:1727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD RES 8
SQ SEQUENCE 8 AA; 919 MW; C82879D5A8569A85 CRC64;

Query Match 31.8%; Score 14; DB 1; Length 8;
Best Local Similarity 50.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SEMF 5
DB 1 SDMY 4

RESULT 4

U1AE_HUMAN STANDARD; PRT; 9 AA.
ID U1AE_HUMAN
AC P31931;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page of liver tissue (Spot 115) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=liver;
RA MEDLINE=9414969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA Tissot J.-D., Balroch A., Appel R.D., Hochstrasser D.F.;
RT "Human liver protein map: update 1993";
RL Electrophoresis 14:1216-1222(1993).
CC -I- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 5.5, ITS MW IS: 11 KDa.
DR SWISS-2DPAGE; P31931; HUMAN.
FT NON_TER 9

SQ SEQUENCE 9 AA; 1096 MW; 477B3B173AE729C7 CRC64;

Query Match 29.5%; Score 13; DB 1; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.1e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 MFL 8
DB 2 LFLXL 6

RESULT 5

E104_LITRU STANDARD; PRT; 5 AA.
ID E104_LITRU
AC P82100;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE E104_LITRU
OS E104_LITRU (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Pelodidae; Hyloria.
OK NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella";
RL Aust. J. Chem. 52:639-645(1999).
CC -I- SUBCELLULAR LOCATION: SECRETED.
KW Amphibian skin; Amidation.
FT MOD RES 5
SQ SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;

Query Match 27.3%; Score 12; DB 1; Length 5;
Best Local Similarity 20.0%; Pred. No. 1.1e+05;
Matches 1; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 FLOIY 9
DB 1 FITVH 5

RESULT 6

UC22_MAIZE STANDARD; PRT; 5 AA.
ID UC22_MAIZE
AC P80628;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of etiolated coleoptile (Spot 474) (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OK NCBI_TaxID=4577;
RN [1]
RP SEQUENCE.
RC TISSUE=Coleoptile;
RA Toulzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,
RA Pernollet J.-C., Zivy M., de Vienne D.;
RT "The maize two dimensional gel protein database: towards an integrated
RT genome analysis program";
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -I- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.1, ITS MW IS: 30.4 KDa.
DR Maize-2DPAGE; P80628; COLEOPTILE.
DR MaizeDB; 123954; -; 1
FT NON_TER 1

```

OX NCBI_TaxID=30608;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP285, JP288, JP289, JP292, JP308, AND JP313;
RX MEDLINE=2118472; PubMed=11286490;
RA Pastori J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224624; AAK70547.1; -.
DR EMBL; AF224625; AAK70551.1; -.
DR EMBL; AF224626; AAK70555.1; -.
DR EMBL; AF224627; AAK70559.1; -.
DR EMBL; AF224628; AAK70563.1; -.
DR EMBL; AF224629; AAK70567.1; -.
KM Mitochondrion.
FT NON TER
SQ SEQUENCE 9 AA; 1160 MW; D5C563636B5045A2 CRC64;

Query Match
Best Local Similarity 34.1%; Score 15; DB 8; Length 9;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
Db 1 YSIV 5

RESULT 7
O94NB1 PRELIMINARY; PRT; 9 AA.
ID O94NB1
AC O94NB1;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Cytochrome oxidase subunit III (Fragment).
GN COI1.
OS Microcebus ravelobensis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
OC Microcebus.
OX NCBI_TaxID=122231;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP289, AND JP301;
RX MEDLINE=2118472; PubMed=11286490;
RA Pastori J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224630; AAK70571.1; -.
DR EMBL; AF224631; AAK70575.1; -.
KM Mitochondrion.
FT NON TER
SQ SEQUENCE 9 AA; 1160 MW; D5C563636B5045A2 CRC64;

Query Match
Best Local Similarity 34.1%; Score 15; DB 8; Length 9;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
Db 1 YSIV 5

RESULT 8
O94NB0 PRELIMINARY; PRT; 9 AA.
ID O94NB0
AC O94NB0;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

```

```

DE Cytochrome oxidase subunit III (Fragment).
GN COI1.
OS Microcebus rufus (brown mouse lemur).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Cheirogaleidae;
OC Microcebus.
OX NCBI_TaxID=122232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP309, JP315, JP316, AND JP317;
RX MEDLINE=2118472; PubMed=11286490;
RA Pastori J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224636; AAK70595.1; -.
DR EMBL; AF224637; AAK70599.1; -.
DR EMBL; AF224638; AAK70603.1; -.
DR EMBL; AF224639; AAK70607.1; -.
KM Mitochondrion.
FT NON TER
SQ SEQUENCE 9 AA; 1160 MW; D5C563636B5045A2 CRC64;

Query Match
Best Local Similarity 34.1%; Score 15; DB 8; Length 9;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
Db 1 YSIV 5

RESULT 9
O94NA9 PRELIMINARY; PRT; 9 AA.
ID O94NA9
AC O94NA9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Cytochrome oxidase subunit III (Fragment).
GN COI1.
OS Daubentonia madagascariensis (Aye-aye).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Daubentonidae;
OC Daubentonia.
OX NCBI_TaxID=31869;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JP119, AND JP120;
RX MEDLINE=2118472; PubMed=11286490;
RA Pastori J., Martin R.D., Ehresmann P., Zimmermann E., Forstner M.R.;
RT "Molecular phylogeny of the lemur family Cheirogaleidae (primates)
  based on mitochondrial DNA sequences.";
RL Mol. Phylogenet. Evol. 19:45-56(2001).
DR EMBL; AF224641; AAK70615.1; -.
DR EMBL; AF224642; AAK70619.1; -.
KM Mitochondrion.
FT NON TER
SQ SEQUENCE 9 AA; 1160 MW; D5C563636B5045A2 CRC64;

Query Match
Best Local Similarity 34.1%; Score 15; DB 8; Length 9;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOQY 9
Db 1 YSIV 5

RESULT 10
O50556

```

ID 05056 PRELIMINARY; PRT; 7 AA.
 AC 05056;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GlyA (Fragment).
 GN GLYA.
 OS Actinobacillus actinomycetemcomitans (Haemophilus
 actinomycetemcomitans)
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Actinobacillus.
 NCBI_TaxId=714;
 RX MEDLINE=96355846; PubMed=8751884;
 RA Kojdodubetz D., Siltzengel J. Jr., Wang B., Phillips L.H., Jacobs C.,
 RA Kraig E.;
 RA "cis Elements and trans factors are both important in strain-specific
 RT regulation of the leukotoxin gene in Actinobacillus
 RT actinomycetemcomitans."
 RL Intec. Immun. 64:3451-3460(1996).
 DR EMBL; U51862; AAB88721.1; -;
 FT NON TER
 SQ SEQUENCE 7 AA; 832 MW; 6DCB42D767340420 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 7;
 Best Local Similarity 50.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 LQTY 9
 DB 3 LPVY 6

RESULT 11
 ID 068485 PRELIMINARY; PRT; 8 AA.
 AC 068485;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Aminoglycoside 3'-O-nucleotidyltransferase (Fragment).
 GN AANA1.
 OS Klebsiella pneumoniae.
 OC Plasmid PQ1000.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 NCBI_TaxId=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1;
 RK MEDLINE=98287600; PubMed=9624504;
 RT "Characterization of the 6'-N-aminoglycoside acetyltransferase gene
 aa(6')-tg from the integration of a natural multiresistance plasmid".
 RL Antimicrob. Agents Chemother. 42:1506-1508(1998).
 DR EMBL; AF047556; AAC25501.1; -;
 KW Plasmid; Transferase.
 FT NON TER
 SQ SEQUENCE 8 AA; 878 MW; EFC1B5A2CH1DD056 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 8;
 Best Local Similarity 42.9%; Pred. No. 6.7e+05;
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ISEMFLO 7
 DB 2 IAEVSTQ 8

RESULT 12
 P83158

ID P83158 PRELIMINARY; PRT; 8 AA.
 AC P83158;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Photosystem I iron-sulfur center (Photosystem I subunit VII) (9 kDa
 DE polypeptide) (PSI-C) (Fragment).
 GN Anabaena sp. (strain 131).
 OS Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 NCBI_TaxId=29412;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Apté S.K., Uhlemann E., Schmid R., Alendort K.;
 RA Submitted (OCT-2001) to the SWISS-PROT data bank.
 RL -1- FUNCTION: APOPROTEIN FOR THE IRON-SULFUR CENTERS PA AND FB OF THE
 CC PHOTOSYSTEM I COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
 CC 'BACTERIAL-TYPE' 4FE4S-FERRDOXIN; PARTIAL.
 DR InterPro; IPRO01450; 4FE4S-FERRDOXIN.
 DR PROSITE; PS00198; 4FE4S-FERRDOXIN; PARTIAL.
 KM Photosynthesis; Photosystem I; Iron-sulfur; 4Fe-4S; Metal-binding.
 FT NON TER
 SQ SEQUENCE 8 AA; 962 MW; C5BB505322D1A1F5 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 8;
 Best Local Similarity 50.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 LQTY 9
 DB 4 VKLY 7

RESULT 13
 ID 09VRD2 PRELIMINARY; PRT; 8 AA.
 AC 09VRD2;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE CG1666 Protein.
 GN CG1666.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RK MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotlier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Jaitani W., Kalush F., Kaipen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusseren D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
 RA Svitek R., Tector C., Turner C., Venter B., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AF03569; AAF50870.1; -;
 DR FlyBase; FBgn040648; CG11666.
 SQ SEQUENCE 8 AA; 1062 MW; ED1B5B044004376 CRC64;

Query Match 29.5%; Score 13; DB 5; Length 8;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFLQI 8
 DB 1 MWIRI 5

RESULT 14

ID Q98YK9 PRELIMINARY; PRT; 8 AA.
 AC Q98YK9;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Truncated pol protein (Fragment).
 GN POL.
 OC Human immunodeficiency virus type 1.
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OK NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=991614;
 RA Schmidt B., Walter H., Moschik G., Paatz C., Werwein M., Schwingel E.,
 RA Korn K.;
 RT "Recovery of HIV-1 pol gene sequences by direct sequencing of
 RT amplification products derived from plasma samples.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF347458; AAK32535.1; -;
 FT NON TER 1
 SQ SEQUENCE 8 AA; 845 MW; 72CDB1DD736CAB8 CRC64;

Query Match 29.5%; Score 13; DB 15; Length 8;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMF 5
 DB 4 LAEAP 8

RESULT 15

ID Q937J8 PRELIMINARY; PRT; 9 AA.
 AC Q937J8;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Putative transposition protein TnIQ (Fragment).
 GN TnIQ.
 OS *Escherichia coli*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OK NCBI_TaxID=562;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH210; TRANSPOSON=TN5057;
 RX MEDLINE=21604134; PubMed=11763242;
 RA Mindlin S.Z., Kholodil G.Y., Gorlenko Z.M., Minakhina S.V.,
 RA Minakhin L.S., Kalyaeva E.S., Kopteva A.V., Petrova M.A.,
 RA Yurieva O.V., Nikiforov V.G.;
 RT "Mercury resistance transposons of Gram-negative environmental
 RT bacteria and their classification.";
 RL Res. Microbiol. 152:811-822(2001).
 DR EMBL; AJ302765; CAC82977.1; -;
 FT NON TER 1
 SQ SEQUENCE 9 AA; 1023 MW; 1B963AB5B7287AA4 CRC64;

Query Match 29.5%; Score 13; DB 2; Length 9;
 Best Local Similarity 40.0%; Pred. No. 6.7e+05;
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMF 5
 DB 5 LSDKF 9

Search completed: May 1, 2003, 23:23:33
 Job time : 30 secs

Thu May 8 16:14:13 2003

us-09-658-621b-26.ra1

Page 1

GenCore version 5.1.4.ps 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:09:26 ; Search time 0.464286 Seconds
(without alignments)
570.332 Million cell updates/sec

Title: US-09-658-621b-26
Perfect score: 44
Sequence: 1 ISEMPLOIX 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/isa/5A COMB.pdb.*
2: /cgn2_6/prodata/1/isa/5B COMB.pdb.*
3: /cgn2_6/prodata/1/isa/5C COMB.pdb.*
4: /cgn2_6/prodata/1/isa/5D COMB.pdb.*
5: /cgn2_6/prodata/1/isa/5E COMB.pdb.*
6: /cgn2_6/prodata/1/isa/5F COMB.pdb.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 37 | 84.1 | 2035 | 2 | US-08-479-537A-2 |
| 2 | 37 | 84.1 | 2035 | 4 | US-09-083-116-2 |
| 3 | 37 | 84.1 | 2035 | 4 | US-09-134-916A-2 |
| 4 | 32 | 72.7 | 2265 | 2 | US-08-149-097D-36 |
| 5 | 32 | 72.7 | 2509 | 2 | US-08-149-097D-35 |
| 6 | 30 | 68.2 | 272 | 4 | US-09-177-165A-27 |
| 7 | 30 | 68.2 | 815 | 4 | US-09-177-165A-24 |
| 8 | 29 | 65.9 | 120 | 1 | US-08-727-118-2 |
| 9 | 29 | 65.9 | 120 | 1 | US-08-248-466B-8 |
| 10 | 29 | 65.9 | 466 | 3 | US-08-526-136-13 |
| 11 | 28 | 63.6 | 351 | 1 | US-08-415-751-41 |
| 12 | 28 | 63.6 | 407 | 3 | US-08-948-597-3 |
| 13 | 28 | 63.6 | 410 | 4 | US-09-348-817A-3 |
| 14 | 28 | 63.6 | 462 | 2 | US-08-477-451-24 |
| 15 | 28 | 63.6 | 487 | 5 | PCT-US95-06556-4 |
| 16 | 28 | 63.6 | 487 | 5 | PCT-US95-06556-4 |
| 17 | 28 | 63.6 | 598 | 2 | US-08-853-659A-53 |
| 18 | 27 | 61.4 | 27 | 2 | US-08-103-170-24 |
| 19 | 27 | 61.4 | 55 | 4 | US-07-741-453A-52 |
| 20 | 27 | 61.4 | 65 | 1 | PCT-US95-04689-7 |
| 21 | 27 | 61.4 | 108 | 4 | US-09-114-001C-5469 |
| 22 | 27 | 61.4 | 320 | 2 | US-08-530-165-7 |
| 23 | 27 | 61.4 | 341 | 1 | US-08-118-270-48 |
| 24 | 27 | 61.4 | 341 | 5 | PCT-US93-08520-48 |
| 25 | 27 | 61.4 | 366 | 4 | US-09-718-832-4 |
| 26 | 27 | 61.4 | 366 | 4 | US-09-718-832-4 |
| 27 | 27 | 61.4 | 366 | 4 | US-09-718-832-4 |

| | | | | | | |
|----|----|------|------|---|-------------------|--------------------|
| 28 | 27 | 61.4 | 366 | 4 | US-09-718-815-4 | Sequence 4, Appl1 |
| 29 | 27 | 61.4 | 369 | 1 | US-07-937-609-19 | Sequence 19, Appl1 |
| 30 | 27 | 61.4 | 369 | 4 | US-08-025-170-19 | Sequence 19, Appl1 |
| 31 | 27 | 61.4 | 384 | 2 | US-08-103-170-10 | Sequence 10, Appl1 |
| 32 | 27 | 61.4 | 387 | 1 | US-08-196-989B-14 | Sequence 14, Appl1 |
| 33 | 27 | 61.4 | 387 | 2 | US-08-760-936-14 | Sequence 14, Appl1 |
| 34 | 27 | 61.4 | 451 | 1 | US-08-417-330A-12 | Sequence 12, Appl1 |
| 35 | 27 | 61.4 | 488 | 1 | US-08-572-225-1 | Sequence 1, Appl1 |
| 36 | 27 | 61.4 | 986 | 4 | US-08-872-757-4 | Sequence 4, Appl1 |
| 37 | 27 | 61.4 | 1296 | 3 | US-08-728-603-15 | Sequence 15, Appl1 |
| 38 | 27 | 61.4 | 1637 | 4 | US-09-718-692-2 | Sequence 2, Appl1 |
| 39 | 27 | 61.4 | 1637 | 4 | US-09-718-852-2 | Sequence 2, Appl1 |
| 40 | 27 | 61.4 | 1637 | 4 | US-09-718-815-2 | Sequence 2, Appl1 |
| 41 | 27 | 61.4 | 1646 | 4 | US-09-535-008-67 | Sequence 67, Appl1 |
| 42 | 27 | 61.4 | 1647 | 4 | US-09-535-008-2 | Sequence 2, Appl1 |
| 43 | 27 | 61.4 | 1649 | 4 | US-09-535-008-75 | Sequence 75, Appl1 |
| 44 | 27 | 61.4 | 1650 | 4 | US-09-535-008-71 | Sequence 71, Appl1 |
| 45 | 27 | 61.4 | 1678 | 4 | US-09-535-008-69 | Sequence 69, Appl1 |

ALIGNMENTS

RESULT 1
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Parent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LARNE, Richard
; APPLICANT: HARVEY, Maria
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teekin, Robin L.
; REGISTRATION NUMBER: 35,020
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537a-2
Query Match 84.1%; Score 37; DB 2; Length 2035;
Best Local Similarity 88.9%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLQY 9
Db 1853 ISEMFLQY 1861
RESULT 2
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
APPLICATION DATA:
PUBLICATION NUMBER: US/09/083,116
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2
Query Match 84.1%; Score 37; DB 4; Length 2035;
Best Local Similarity 88.9%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLQY 9
Db 1853 ISEMFLQY 1861
RESULT 3
US-09-134-916A-2
Sequence 2, Application US/09134916A
Patent No. 6328956
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre

Thu May 8 16:14:13 2003

us-09-658-621b-26.ra1

Page 3

APPLICANT: KIENV, Marie-Paule
APPLICANT: LATHÉ, Richard
APPLICANT: HARETVENI, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,916A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: PR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teakin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2
Query Match 84.1%; Score 37; DB 4; Length 2035;
Best Local Similarity 88.9%; Pred. No. 167
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ISEMPL01Y 9
DB 1853 ISEMPL01Y 1861
RESULT 4
US-08-149-097D-36
Sequence 36 Application US/08149097D
Patent No 5874236
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,097D
FILING DATE: 05-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA: US 07/482,384
FILING DATE: 20-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:

Thu May 8 16:14:13 2003

us-09-658-621b-26.rat

Page 4

```

; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "Alpha1-2 subunit of
; US-08-149-097D-36
; OTHER INFORMATION: human calcium channel"

Query Match 72.7%; Score 32; DB 2; Length 2265;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMPL01Y 9
DB 530 MSEMPLKMY 538

RESULT 5
US-08-149-097D-35
; Sequence 35, Application US/08149097D
; Patent No. 5874236
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,097D
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US92/06903
; FILING DATE: 14-AUG-1992
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2509 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "Alpha1-1 subunit of
; US-09-149-097D-35
; OTHER INFORMATION: human calcium channel"

Query Match 72.7%; Score 32; DB 2; Length 2509;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMPL01Y 9
DB 530 MSEMPLKMY 538

RESULT 6
US-09-177-165A-27
; Sequence 27, Application US/09177165A
; Patent No. 6426205
; GENERAL INFORMATION:
; APPLICANT: Tyers, Mike
; APPLICANT: Williams, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; TITLE OF INVENTION: DEPENDENT PROTEOLYSIS
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/09/177,165A
; FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
```

```

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-27

Query Match      68.2%; Score 30; DB 4; Length 272;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
   |||||
Db 98 SETFLQFY 105

RESULT 7
US-09-177-165A-24
; Sequence 24, Application US/09177165A
; Patent No. 6426205
; GENERAL INFORMATION:
; APPLICANT: Tyers, Mike
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; FILE REFERENCE: 11757,10USU1
; CURRENT APPLICATION NUMBER: US/09/177,165A
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 24
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-24

Query Match      68.2%; Score 30; DB 4; Length 815;
Best Local Similarity 75.0%; Pred. No. 19e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SEMFLQIY 9
   |||||
Db 106 SETFLQFY 113

RESULT 8
US-08-727-118-2
; Sequence 2, Application US/08727118
; Patent No. 5928940
; GENERAL INFORMATION:
; APPLICANT: MIYAZONO, KOHEI
; APPLICANT: SAMBATH, KUBER T.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: NOVEL MORPHOGEN-RESPONSIVE SIGNAL
; TITLE OF INVENTION: TRANSDUCER AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CREATIVE BIOMOLECULES, INC.
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/727,118
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: VITO PHD, CHRISTINE C.
; REGISTRATION NUMBER: 39,061
; REFERENCE/DOCKET NUMBER: CRP-121 [2034/91]
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-727-118-2

Query Match      65.9%; Score 29; DB 2; Length 120;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
   |||||
Db 96 LQSEFLQIY 104

RESULT 9
US-08-248-466B-8
; Sequence 8, Application US/08248466B
; Patent No. 5629182
; GENERAL INFORMATION:
; APPLICANT: CHOPIN, MARIE-CHRISTINE
; APPLICANT: CLUZEL, PIERRE-JEAN
; TITLE OF INVENTION: CLONING OF DNA FRAGMENTS ENCODING A
; TITLE OF INVENTION: MECHANISM FOR RESISTANCE TO BACTERIOPHAGES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,466B
; FILING DATE: 24-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,959
; FILING DATE: 15-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/11381
; FILING DATE: 14-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5629182man F.
; REGISTRATION/DOCKET NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 384-033-0 PCT FWC CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
```

MOLECULE TYPE: protein

US-08-248-466B-8
Query Match 65.9%; Score 29; DB 1; Length 224;
Best Local Similarity 55.6%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
Db 143 ISEMFLQY 151

RESULT 10

US-08-526-136-13
Sequence 13, Application US/08526136
Patent No. 6107089
GENERAL INFORMATION:
APPLICANT: Towle, Christine A. et al.
TITLE OF INVENTION: ANNEXIN XI
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 507 or 558X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526.136
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214.036
FILING DATE:
APPLICATION NUMBER: 07/837.775
FILING DATE: February 13, 1992
APPLICATION NUMBER: 07/764.465
FILING DATE: September 23, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/099001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TRFAX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 466
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-526-136-13

Query Match 65.9%; Score 29; DB 3; Length 466;
Best Local Similarity 55.6%; Pred. No. 1,7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
Db 432 IKQMFQY 440

RESULT 11

US-08-415-751-41
Sequence 41, Application US/08415751
Patent No. 5643772
GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI

TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS AND KIT
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415.751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071.880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891.301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolazalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
FEATURE:
NAME/KEY: Positions coded by nonsense codons are
NAME/KEY: Identified as Xaa.

US-08-415-751-41
Query Match 63.6%; Score 28; DB 1; Length 351;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
Db 7 IASLYQY 15

RESULT 12

US-08-948-997-3
Sequence 3, Application US/08948997
Patent No. 6008020
GENERAL INFORMATION:
APPLICANT: HASTINGS, GREGG
APPLICANT: COLEMAN, TIM
APPLICANT: LAWRENCE, DANIEL
TITLE OF INVENTION: BRAIN-ASSOCIATED INHIBITOR OF
TITLE OF INVENTION: TISSUE-TYPE PLASMINOGEN ACTIVATOR
NUMBER OF SEQUENCES: 17

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
;; STREET: 9410 KEY WEST AVENUE
;; CITY: ROCKVILLE
;; STATE: MD
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/948,997
;; FILING DATE: Oct-10-97
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: A. ANDERS BROOKES
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PF336
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 407 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-948-997-3

Query Match 63.6%; Score 28; DB 3; Length 407;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
DB 118 VSEKFLQ 125

RESULT 13
US-09-348-817A-3
; Sequence 3, Application US/09348817A
; Patent No. 6191260
; GENERAL INFORMATION:
; APPLICANT: Hastings et al.
; TITLE OF INVENTION: Brain-Associated Inhibitor of Tissue-Type Plasminogen
; TITLE OF INVENTION: Activator
; FILE REFERENCE: PF336D1
; CURRENT APPLICATION NUMBER: US/09/348,817A
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 08/948,997
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: 60/028,117
; PRIOR FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-348-817A-3

Query Match 63.6%; Score 28; DB 4; Length 410;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
DB 120 VSEKFLQ 127

RESULT 14
US-08-477-451-24
; Sequence 24, Application US/08477451
; Patent No. 592865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;; US-08-477-451-24

Query Match 63.6%; Score 28; DB 2; Length 462;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 EMFLQIY 9
DB 229 QIFLOIY 235

RESULT 15
US-08-249-112-4
; Sequence 4, Application US/08249112
; Patent No. 5527703
; GENERAL INFORMATION:
; APPLICANT: Cully, Doris F.
; APPLICANT: Arena, Joseph P.
; APPLICANT: Liu, Ken K.
; APPLICANT: Vassiliadis, Demetrios
; TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
; TITLE OF INVENTION: CHANNELS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wallen, John W.
; STREET: 126 E. Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/249,112
: FILING DATE: 25-MAY-1994
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
:   NAME: Wallen, John W.
:   REGISTRATION NUMBER: 35,403
:   REFERENCE/DOCKET NUMBER: 19194
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (908) 594-3905
:   TELEFAX: (908) 594-4720
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 487 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-249-112-4

```

```

Query Match      63.6%; Score 28; DB 1; Length 487;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 ISEMPLQIY 9
Db 140 MEMPLR1Y 148

```

Search completed: May 1, 2003, 23:14:27
 Job time : 1.46429 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:30 ; Search time 0.321423 Seconds

(without alignments)
1161.337 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44
Sequence: 1 ISENFQIY 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 44 | 100.0 | 1255 | 1 MUC1_HUMAN | P15941 h mucin 1 P |
| 2 | 35 | 79.5 | 475 | 1 MUC1_HYLLA | Q29435 hylobates 1 |
| 3 | 33 | 75.0 | 2223 | 1 CCAB_DISOM | P56699 discopoda o |
| 4 | 32 | 72.7 | 1848 | 1 CCAB_DROME | P91645 drosophila |
| 5 | 32 | 72.7 | 2164 | 1 CCAB_MOUSE | P97445 mus musculu |
| 6 | 32 | 72.7 | 2212 | 1 CCAB_RAT | P54282 rattus norv |
| 7 | 32 | 72.7 | 2424 | 1 CCAB_RABIT | P27884 coryctolagus |
| 8 | 32 | 72.7 | 2505 | 1 CCAB_HUMAN | O00555 homo sapien |
| 9 | 31 | 70.5 | 144 | 1 SODM_PALVU | P28765 palinurus v |
| 10 | 31 | 70.5 | 366 | 1 VGLM_HSVSA | O01017 herpesvirus |
| 11 | 31 | 70.5 | 491 | 1 VGB4_MYCTU | O53209 mycobacteri |
| 12 | 31 | 70.5 | 604 | 1 YB2A_ECOLI | P33913 escherichia |
| 13 | 31 | 70.5 | 896 | 1 APCE_SYNY4 | Q02907 synechocyst |
| 14 | 31 | 70.5 | 1071 | 1 PPI6_YEAST | P15938 saccharomyc |
| 15 | 31 | 70.5 | 1630 | 1 MSP1_PLAFK | P04933 plasmodium |
| 16 | 31 | 70.5 | 1639 | 1 MSP1_PLAFW | Q10105 schizosach |
| 17 | 31 | 70.5 | 2670 | 1 YAO5_SCHPO | Q9YCS9 drosophila |
| 18 | 30 | 68.2 | 387 | 1 O94A_DROME | O02496 mus musculu |
| 19 | 30 | 68.2 | 630 | 1 MUC1_MOUSE | P35939 saccharomyc |
| 20 | 30 | 68.2 | 772 | 1 PM1E_YEAST | Q12018 saccharomyc |
| 21 | 30 | 68.2 | 815 | 1 CCS3_YEAST | Q12751 saccharomyc |
| 22 | 30 | 68.2 | 981 | 1 YV48_YEAST | Q12750 porcine rot |
| 23 | 30 | 68.2 | 1082 | 1 RPOB_RORPC | Q9L106 nephroselm |
| 24 | 30 | 68.2 | 1109 | 1 RPOB_NEPOL | O92999 chlamydia p |
| 25 | 30 | 68.2 | 1393 | 1 RPOC_CHLPM | O92999 chlamydia m |
| 26 | 30 | 68.2 | 1396 | 1 RPOC_CHLPM | O92999 chlamydia t |
| 27 | 30 | 68.2 | 1396 | 1 RPOC_CHLPM | P31114 bacillus su |
| 28 | 29 | 65.9 | 348 | 1 HBP2_BACSU | O07076 mus musculu |
| 29 | 29 | 65.9 | 463 | 1 ANK7_MOUSE | P20076 mus musculu |
| 30 | 29 | 65.9 | 466 | 1 ANK7_HUMAN | P20076 mus musculu |
| 31 | 29 | 65.9 | 527 | 1 SYK_CHLPM | O92999 chlamydia p |
| 32 | 29 | 65.9 | 548 | 1 ALGB_YEAST | P53954 saccharomyc |
| 33 | 29 | 65.9 | 558 | 1 ATVA_SYNY3 | P73866 synechocyst |

| | | | | | |
|----|----|------|------|--------------|--------------------|
| 34 | 29 | 65.9 | 558 | 1 YNT3_YEAST | P53870 saccharomyc |
| 35 | 29 | 65.9 | 598 | 1 LEPA_MYCPN | P75498 mycoplasma |
| 36 | 29 | 65.9 | 673 | 1 UVRB_BORBU | O51776 borrelia bu |
| 37 | 29 | 65.9 | 806 | 1 SRCA_MYCGE | P47318 mycoplasma |
| 38 | 29 | 65.9 | 808 | 1 SRCA_MYCPN | P75559 mycoplasma |
| 39 | 29 | 65.9 | 1058 | 1 U202_ARATH | O92144 arabidopsis |
| 40 | 29 | 65.9 | 1071 | 1 TR1_THESC | P66086 thermoplas |
| 41 | 29 | 65.9 | 1262 | 1 RRP1_P1ZHT | P26676 human para |
| 42 | 28 | 63.6 | 112 | 1 Y122_CABEL | P03935 caenorhabd1 |
| 43 | 28 | 63.6 | 117 | 1 Y627_METJA | O58044 methanococ |
| 44 | 28 | 63.6 | 188 | 1 Y101_UREPA | O9PR43 ureaplasma |
| 45 | 28 | 63.6 | 299 | 1 YX06_CABEL | Q11113 caenorhabd1 |

ALIGNMENTS

| | |
|--|-------------------------|
| RESULT 1 | |
| ID MUC1_HUMAN | STANDARD: PRT; 1255 AA. |
| AC P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4U2; | |
| AC Q16615; Q14876; Q9UE75; Q9UE76; Q9U0L1; Q9BXA4; | |
| DT 01-JAN-1990 (Rel. 13, Created) | |
| DT 01-APR-1990 (Rel. 14, Last sequence update) | |
| DT 15-JUN-2002 (Rel. 41, Last annotation update) | |
| DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT) | |
| DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin) | |
| DE (Tumor-associated epithelial membrane antigen) (EMA) (H2JAG) (Peanut- | |
| DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen | |
| DE DP3) (CD227 antigen). | |
| GN MUC1. | |
| OS Homo sapiens (Human). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| OX NCBI_TaxID=9606; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. (ISOFORM A). | |
| RC TISSUE=pancreas; | |
| RX MEDLINE=90368716; PubMed=2394722; | |
| RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.; | |
| RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA."; | |
| RL J. Biol. Chem. 265:15294-15299(1990). | |
| RN [2] | |
| RP SEQUENCE FROM N.A. (ISOFORM A AND B). | |
| RX MEDLINE=9020794; PubMed=2318825; | |
| RA Ligenberg M.J., Vos H.L., Gennissen A.M.C., Hilkens J.; | |
| RT "Episialin, a carcinoma-associated mucin, is generated by a | |
| RT polymorphic gene encoding splice variants with alternative amino | |
| RT termini."; | |
| RL J. Biol. Chem. 265:5573-5578(1990). | |
| RN [3] | |
| RP SEQUENCE FROM N.A. (ISOFORM A). | |
| RC TISSUE=Breast carcinoma; | |
| RX MEDLINE=90368715; PubMed=1697589; | |
| RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.; | |
| RT "Molecular cloning and expression of human tumor-associated | |
| RT polymorphic epithelial mucin."; | |
| RL J. Biol. Chem. 265:15286-15293(1990). | |
| RN [4] | |
| RP SEQUENCE FROM N.A. (ISOFORM A). | |
| RX MEDLINE=91097524; PubMed=2268309; | |
| RA Lancaster C.A., Peat N., Duhig T., Wilson D.; | |
| RT Taylor-Papadimitriou J., Gendler S.J.; | |
| RT "Structure and expression of the human polymorphic epithelial mucin | |
| RT gene: an expressed VNTR unit."; | |
| RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990). | |
| RN [5] | |
| RP SEQUENCE FROM N.A. (ISOFORM SEC). | |
| RC TISSUE=Breast carcinoma; | |
| RX MEDLINE=90276413; PubMed=2351132; | |
| RA Wreschner D.H., Harevent M., Tsarfaty I., Smorodinsky N., Horev J.; | |
| RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.; | |

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Harevuani M., Tsarfaty I., Zaretsky J., Kockes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lache R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=9103045; PubMed=1688329;
RA Tsarfaty I., Harevuani M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lache R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM Y).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORMS X; Y AND Z).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RP SEQUENCE FROM N.A. (ISOFORM Y).
RX Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A. (ISOFORM S).
RC TISSUE=Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM B).
RX MEDLINE=90086473; PubMed=2597151;
RA Abe M., Siddiqui J., Kute D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM B).
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";

RL Int. J. Cancer 66:55-59(1996).
RN [15]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [16]
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS C AND D).
RC TISSUE=Breast carcinoma;
RX Bihwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=97460054; PubMed=9312074;
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
RA Hanisch F.-G.;
RT "Localization of O-glycosylation sites on glycopeptide fragments from
RT lactation-associated MUC1. All putative sites within the tandem
RT repeat are glycosylation targets in vivo.";
RL J. Biol. Chem. 272:24780-24793(1997).
RN [18]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=99303572; PubMed=10373415;
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
RA Hanisch F.-G.;
RT "High density O-glycosylation on tandem repeat peptide from secretory
RT MUC1 of T47D breast cancer cells.";
RL J. Biol. Chem. 274:18165-18172(1999).
RN [19]
RP POLYMORPHISM WITHIN THE REPEAT.
RX MEDLINE=21359366; PubMed=11350974;
RA Engelmann K., Baldus S.E., Hanisch F.-G.;
RT "Identification and topology of variant sequences within individual
RT repeat domains of the human epithelial tumor mucin MUC1.";
RL J. Biol. Chem. 276:27764-27769(2001).
RN [20]
RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RX MEDLINE=99211485; PubMed=10197628;
RA Baruch A., Hartmann M.-L., Yoeli M., Aderech Y., Greenstein S.,
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA Wreschner D.H.;
RT "The breast cancer-associated MUC1 gene generates both a receptor and
RT its cognate binding protein.";
RL Cancer Res. 59:1552-1561(1999).
RN [21]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX MEDLINE=21240104; PubMed=11341784;
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RA Harris A.;
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN [22]
RP CHARACTERIZATION.
RX MEDLINE=2183452; PubMed=11847293;
RA Wreschner D.H., McShuckin M.A., Williams S.J., Baruch A., Yoeli M.,
RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA Stacey M., Lin H.-H., Gordon S.;
RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
RT cleavage of membrane-associated mucin proteins.";
RL Protein Sci. 11:698-706(2002).
RN [23]
RP PHOSPHORYLATION.
RX MEDLINE=95080414; PubMed=7988707;
RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
RT Cytokine receptor-like molecules.";
RL FEBS Lett. 356:130-136(1994).
RN CC
CC -I- FUNCTION: May play a role in adhesive functions and in cell-cell
interactions, metastasis and signaling. May provide a protective

Query Match 100.0%; Score 44; DB 1; Length 1255;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQTY 9
 DB 1073 ISEMFLQTY 1081

RESULT 2
 MUC1_HYLLA STANDARD; PRT; 475 AA.
 ID MUC1_HYLLA
 AC Q29435;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mucin 1 precursor (MOC-1).
 GN MUC1.
 OS Hylobates lar (Common gibbon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
 OC NCBI_TaxID=9580;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96351712; PubMed=8747930;
 RA Spiller A.P., Dahlg T., Chilton B.S., Gendler S.J.;
 RT "Analysis of mammalian MUC1 genes reveals potential functionally
 important domains.";
 RL Mamm. Genome 6:885-888(1995).
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
 CYTOSKELETON (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: IMA589; AAA69965.1; -;
 DR EMBL: I41625; AAA69918.1; -;
 DR EMBL: I41624; AAA69918.1; JOINED.
 DR InterPro: IPR000082; SEA_domain.
 DR Pfam: PF01390; SEA; 1.
 DR SMART: SM00200; SEA; 1.
 DR PROSITE: PS50024; SEA; 1.
 KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
 Repeat.
 KM Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 475 MUCIN 1.
 FT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 381 401 POTENTIAL.
 FT DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 102 121 1.
 FT REPEAT 122 141 2.
 FT REPEAT 142 161 3.
 FT REPEAT 162 181 4.
 FT DOMAIN 254 371 SEA.
 SQ SEQUENCE 475 AA; 49371 MW; D7A699D6D68C6622 CRC64;
 Query Match 79.5%; Score 35; DB 1; Length 475;
 Best Local Similarity 77.8%; Pred. No. 4.5;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 3
 CCAE DISOM STANDARD; PRT; 2223 AA.
 ID CCAE DISOM
 AC p56699;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable voltage-dependent R-type calcium channel alpha-1E subunit
 (DOE-1).
 OS Discosoma ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hyphosqualea; Pristiogalea; Batoidae;
 OC Torpediniformes; Narcinidae; Narcinidae; Discosoma.
 OC NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Electric lobe;
 RX MEDLINE=93248175; PubMed=7683405;
 RA Horne W.A., Ellinger P.T., Inman I., Zhou M., Tsien R.W., Schwarz T.L.;
 RT "Molecular diversity of Ca2+ channel alpha 1 subunits from the marine
 ray Discosoma ommata.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3787-3791(1993).
 CC -1- FUNCTION: THE ISOFORM ALPHA-1E GIVES RISE TO THE "HIGH-VOLTAGE
 CURRENTS. R-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE
 ACTIVATED" (HVA) GROUP (BY SIMILARITY).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN THE FOREBRAIN THAN IN
 THE ELECTRIC LOBE.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L12531; -; NOT ANNOTATED CDS.
 DR InterPro: IPR001682; Ca/Na_pore.
 DR InterPro: IPR002077; Ca channel.
 DR InterPro: IPR002111; Ca channel_TripL.
 DR InterPro: IPR000636; M-channel_mng.
 DR Pfam: PF00520; Ion trans; 4.
 DR PRINTS: PRO0167; CACHANNEL.
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Calcium channel; Glycoprotein; Repeat; Multigene family;
 KM Calcium-binding; Phosphorylation.
 FT REPEAT 89 366 I.
 FT REPEAT 478 720 II.
 FT REPEAT 1092 1377 III.
 FT REPEAT 1414 1666 IV.
 FT DOMAIN 1 102 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 103 121 S1 OF REPEAT I (POTENTIAL).
 FT DOMAIN 122 139 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 139 159 S2 OF REPEAT I (POTENTIAL).
 FT DOMAIN 160 171 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 172 189 S3 OF REPEAT I (POTENTIAL).

| | | | | |
|----|----------|-----------------------|------------------|--|
| FT | DOMAIN | 190 | 194 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 195 | 213 | S4 OF REPEAT I (POTENTIAL). |
| FT | TRANSMEM | 214 | 232 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 233 | 252 | S5 OF REPEAT I (POTENTIAL). |
| FT | TRANSMEM | 253 | 338 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 339 | 363 | S6 OF REPEAT I (POTENTIAL). |
| FT | TRANSMEM | 364 | 490 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 491 | 509 | S1 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 510 | 524 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 525 | 544 | S2 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 545 | 552 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 553 | 571 | S3 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 572 | 581 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 582 | 600 | S4 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 601 | 619 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 620 | 639 | S5 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 640 | 692 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 693 | 717 | S6 OF REPEAT II (POTENTIAL). |
| FT | TRANSMEM | 718 | 1105 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1106 | 1124 | S1 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1125 | 1140 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1141 | 1160 | S2 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1161 | 1172 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1173 | 1191 | S3 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1192 | 1205 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1206 | 1224 | S4 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1225 | 1243 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1244 | 1263 | S5 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1264 | 1349 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1350 | 1374 | S6 OF REPEAT III (POTENTIAL). |
| FT | TRANSMEM | 1375 | 1429 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1430 | 1448 | S1 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1449 | 1463 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1464 | 1483 | S2 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1484 | 1491 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1492 | 1510 | S3 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1511 | 1519 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1520 | 1537 | S4 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1538 | 1558 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 1559 | 1577 | S5 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1578 | 1638 | EXTRACELLULAR (POTENTIAL). |
| FT | TRANSMEM | 1639 | 1654 | S6 OF REPEAT IV (POTENTIAL). |
| FT | TRANSMEM | 1654 | 2223 | CYTOPLASMIC (POTENTIAL). |
| FT | TRANSMEM | 2223 | 33 | POLY-GLU. |
| FT | TRANSMEM | 729 | 735 | POLY-SER. |
| FT | TRANSMEM | 1064 | 1068 | BINDING TO THE BETA SUBUNIT (BY |
| FT | TRANSMEM | 386 | 403 | SIMILARITY). |
| FT | TRANSMEM | 321 | 321 | CALCIUM ION SELECTIVITY AND PERMEABILITY |
| FT | TRANSMEM | 671 | 671 | (BY SIMILARITY). |
| FT | TRANSMEM | 1323 | 1323 | CALCIUM ION SELECTIVITY AND PERMEABILITY |
| FT | TRANSMEM | 1611 | 1611 | (BY SIMILARITY). |
| FT | TRANSMEM | 439 | 450 | CALCIUM ION SELECTIVITY AND PERMEABILITY |
| FT | TRANSMEM | 1674 | 1674 | (BY SIMILARITY). |
| FT | TRANSMEM | 1692 | 1703 | CALCIUM ION SELECTIVITY AND PERMEABILITY |
| FT | TRANSMEM | 287 | 287 | (BY SIMILARITY). |
| FT | TRANSMEM | 1300 | 1300 | N-LINKED (GLCNAc. . .) (POTENTIAL). |
| FT | TRANSMEM | 1519 | 1519 | N-LINKED (GLCNAc. . .) (POTENTIAL). |
| FT | TRANSMEM | 2223 | 2223 | N-LINKED (GLCNAc. . .) (POTENTIAL). |
| FT | TRANSMEM | 251827 | 251827 | N-LINKED (GLCNAc. . .) (POTENTIAL). |
| FT | TRANSMEM | 59722DC03B1CFC5B | 59722DC03B1CFC5B | CRC64; |
| FT | TRANSMEM | 75.0%; | Score 33; | DB 1; Length 2223; |
| FT | TRANSMEM | Best Local Similarity | 75.0%; | Pred. No. 65; |
| FT | TRANSMEM | Matches 6; | Conservative 2; | Mismatches 0; Indels 0; Gaps 0; |

Qy 2 SEMFLQTY 9
 Db 535 SEMFLQTY 542

| | | | |
|----------|---|-----------|---------------|
| RESULT 4 | CCAA_DROME | STANDARD; | PRT; 1848 AA. |
| ID | CCAA_DROME | STANDARD; | PRT; 1848 AA. |
| AC | P9145; C01173; C01714; C09YR8; | | |
| DT | 15-JUN-1998 (Rel. 36, Created) | | |
| DT | 15-JUN-2002 (Rel. 41, Last sequence update) | | |
| DE | Voltage-dependent calcium channel type A alpha-1 subunit (Cacophony protein) (NIGHTBLIND A protein) (No-on-transient B protein) (DMCA1A). | | |
| DE | CAC OR NBA OR NONB OR CG1522. | | |
| GN | Drosophila melanogaster (Fruit fly). | | |
| OS | Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; | | |
| OC | Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | |
| OC | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | |
| OX | NCBI_Taxid:72271 | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=97141514; PubMed=8987815; | | |
| RX | Smith L.A., Wang X.J., Peixoto A.A., Neumann E.K., Hall L.M., | | |
| RT | "A Drosophila calcium channel alpha subunit gene maps to a genetic | | |
| RT | locus associated with behavioral and visual defects." | | |
| RT | J. Neurosci. 16:7868-7879 (1996). | | |
| RL | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=Berkeley; | | |
| RC | MEDLINE=20196006; PubMed=10731132; | | |
| RX | Adams W.D., Gelnhaker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D., | | |
| RA | Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abriil J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D., | | |
| RA | Ballem R.M., Baer A., Bakendell J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Boltskov S., | | |
| RA | Borckova D., Botchan M.R., Bouck J., Brokstein P., Brotlier F., | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu L.B., Davies A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K.E., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., | | |
| RA | Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K., | | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honick J., | | |
| RA | Hoskins R., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C., | | |
| RA | Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.R., Ketchum K.A., | | |
| RA | Kamel B.E., Kodira C.D., Krat C., Kravitz S., Kuip D., Lai Z., | | |
| RA | Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., | | |
| RA | Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., | | |
| RA | Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., | | |
| RA | Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., | | |
| RA | Nelson D.R., Nelson X.A., Nixon K., Nuskern D.R., Pauley J.M., | | |
| RA | Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., | | |
| RA | Reichert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., | | |
| RA | Shue B.C., Sideris-Kiamos I., Simpson M., Stupski M.P., Smith T., | | |
| RA | Spiet E., Spradling A.C., Stapleton M., Strong R., Sun E., | | |
| RA | Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J., | | |
| RA | Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., | | |
| RA | Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., | | |
| RA | Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., | | |
| RA | Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; | | |
| RT | "The genome sequence of Drosophila melanogaster." | | |
| RL | Science 287:2185-2195 (2000). | | |
| RL | [3] | | |
| RP | PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING. | | |
| RP | MEDLINE=97247730; PubMed=9039853; | | |
| RX | Peixoto A.A., Smith L.A., Hall J.C.; | | |
| RT | "Genomic organization and evolution of alternative exons in a | | |
| RT | Drosophila calcium channel gene." | | |
| RL | Genetics 145:1003-1013 (1997). | | |
| RL | [4] | | |

RP VARIANT CAC-S.
 RA MEDLINE=9835104; PubMed=949530;
 RX Smith L.A., Pelozo A.A., Kramer E.M., Villella A., Hall J.C.;
 RT "Courtship and visual defects of cacophony mutants impair functional
 complexity of a calcium-channel alpha subunit in *Drosophila*."; *Genetics* 149:1407-1426(1998).
 RL
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
 ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CONTRACTION, NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL
 MOTILITY, CELL DIVISION AND CELL DEATH (BY SIMILARITY). PROBABLY
 ENCODES A DIHYDROPYRIDINE-INSENSITIVE CURRENT. VITAL FOR SURVIVAL
 TO ADULTHOOD.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 alternative splicing. At least 2 regions (184 AND 1/11) undergo
 alternative splicing. The total number of isoforms is currently
 not known.
 CC -1- TISSUE SPECIFICITY: EXPRESSED WIDELY IN THE EMBRYONIC NERVOUS
 SYSTEM.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS IN THE FIRST LARVAL INSTAR,
 MIDPUPAL, AND LATE PUPAL STAGES. IN LATE-STAGE EMBRYOS, IT IS
 EXPRESSED PREFERENTIALLY IN THE NERVOUS SYSTEM.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 FAMILY.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL, U55776; AAC7406.1; -;
 DR EMBL, AE003487; AAF48120.1; -;
 DR EMBL, U88664; AAB53271.1; -;
 DR EMBL, U88665; AAB53272.1; -;
 DR FLYBASE; FBGN0005563; cac.
 DR InterPro; IPR001682; Ca/Na_pore.
 DR InterPro; IPR002077; Ca_channel.
 DR InterPro; IPR002048; Ca_channel_TripL.
 DR InterPro; IPR000636; EF-hand.
 DR InterPro; IPR003915; PKD_2.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF00520; ion_trans; 4.
 DR PRINTS; PR00167; CACHANNEL.
 DR PRINTS; PR01433; POLYCYSTRIN2.
 DR SMART; SM00054; Efh; 1.
 KM Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KM Calcium channel; Glycoprotein; Repeat; Multigene family;
 KM Calcium-binding; Phosphorylation; Alternative splicing;
 KM Developmental protein.
 FT REPEAT 25 316 1.
 FT REPEAT 316 670 11.
 FT REPEAT 427 670 11.
 FT REPEAT 762 1049 111.
 FT REPEAT 1086 1344 111.
 FT DOMAIN 1 38
 FT TRANSMEM 39 57 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 58 75 S1 OF REPEAT I (POTENTIAL).
 FT TRANSMEM 76 95 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 96 107 S2 OF REPEAT I (POTENTIAL).
 FT TRANSMEM 108 128 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 129 133 S3 OF REPEAT I (POTENTIAL).
 FT TRANSMEM 134 152 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 153 171 S4 OF REPEAT I (POTENTIAL).
 FT TRANSMEM 172 191 S5 OF REPEAT I (POTENTIAL).

FT DOMAIN 192 288 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 289 313 S6 OF REPEAT I (POTENTIAL).
 FT DOMAIN 314 441 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 442 460 S1 OF REPEAT II (POTENTIAL).
 FT DOMAIN 461 475 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 476 495 S2 OF REPEAT II (POTENTIAL).
 FT DOMAIN 496 504 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 504 522 S3 OF REPEAT II (POTENTIAL).
 FT DOMAIN 523 531 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 532 550 S4 OF REPEAT II (POTENTIAL).
 FT DOMAIN 551 569 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 570 589 S5 OF REPEAT II (POTENTIAL).
 FT DOMAIN 590 642 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 643 667 S6 OF REPEAT II (POTENTIAL).
 FT DOMAIN 668 767 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 768 786 S1 OF REPEAT III (POTENTIAL).
 FT DOMAIN 787 802 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 803 822 S2 OF REPEAT III (POTENTIAL).
 FT DOMAIN 823 834 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 835 853 S3 OF REPEAT III (POTENTIAL).
 FT DOMAIN 854 866 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 867 885 S4 OF REPEAT III (POTENTIAL).
 FT DOMAIN 886 904 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 905 924 S5 OF REPEAT III (POTENTIAL).
 FT DOMAIN 925 1013 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1014 1038 S6 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1039 1093 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1094 1122 S1 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1123 1127 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1128 1147 S2 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1148 1155 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1156 1174 S3 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1175 1181 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1182 1200 S4 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1201 1219 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1220 1239 S5 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1240 1305 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1306 1330 S6 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1331 1848 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1332 1848 POLY-GLU.
 FT DOMAIN 396 402 POLY-GLU.
 FT TRANSMEM 403 422 ARG-RICH.
 FT DOMAIN 423 735 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT TRANSMEM 736 752 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT DOMAIN 753 791 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT TRANSMEM 792 821 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT SITE 822 621 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT SITE 823 989 CALCULON ION SELECTIVITY AND PERMEABILITY.
 FT SITE 990 1278 TO PHENYLALANINES (BY SIMILARITY).
 FT SITE 1279 1345 PHOSPHORYLATION (BY PKA) (POTENTIAL).
 FT MOD RES 1346 1345 BY SIMILARITY.
 FT CA BIND 1346 1374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1375 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 236 865 FMTOYDGEVDLRLAIVRLPKLVSGIP -> AMTI
 FT VARSPLIC 121 153 FAENIDVIDLRLMRSFRLVRLPKLVSRIP (IN ISOFORM
 WITH 184A).
 FT VARSPLIC 315 352 EFSNRNRVRRMRPFQKRFAMQTFAMVSLDITQ ->

Query Match Score 32; DB 1; Length 1848;
 Best Local Similarity 55.6%; Pred. No. 89;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLOIY 9
 Db 485 MSEMIRKMY 493
 RESULT 5
 ID CCAA MOUSE STANDARD; PRT; 2164 AA.
 AC P97445
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Voltage-dependent P/Q-type calcium channel α -1A subunit (Calcium
DE channel, L type, α -1 polypeptide isoform 4) (Brain calcium
DE channel 1) (B1).
GN CACNA1A OR CACNA1A4 OR CCHAI1A OR CACNA4 OR CACNA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT TG LEU-601.
RC STRATIN=DR/2U;
RK MEDLINE=97083572; PubMed=8929530;
RA Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,
RA Hawkes R., Finkbein W.N., Copeland N.G., Jenkins N.A.;
RT "Absence epilepsy in tottering mutant mice is associated with calcium
RT channel defects".
RL Cell 87:607-617(1996).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/O-TYPE CALCIUM
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN THE
CC CEREBELLUM, OLFACTORY BULB, CEREBRAL CORTEX, HIPPOCAMPUS, AND
CC INFERIOR COLICULUS. IN THE HIPPOCAMPUS, EXPRESSION OCCURS IN
CC PYRAMIDAL AND GRANULE NEURONS, AS WELL AS IN INTERNEURONS.
CC PURKINJE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC. THE Q-TYPE BEING
CC A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- DISEASE: DEFECTS IN CACNA1A ARE THE CAUSE OF A DELAYED-ONSET,
CC RECESSIVE NEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,
CC RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE
CC SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN
CC HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)) THAT IS
CC CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST MEANING,
CC BUT NO MOTOR SEIZURES; AND ROLLING (NGOTA (TG(ROL))), THAT PRESENTS
CC AN INTERMEDIARY PHENOTYPE. THE ATAXIA BEING SOMEWHAT MORE SEVERE
CC THAT WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION
CC OF CEREBELLAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF
CC MUTANTS.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U76716; AAC52940.1; -;
DR MGD; MG1:109482; Cacnala.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR001662; Ca/Na_pore.

DR InterPro; IPR002077; Ca_channel.
DR InterPro; IPR002111; Ca_channel_trpl.
DR InterPro; IPR000636; M+channel_nig.
DR Pfam; PF00520; Ion_trans_4.
DR PRINTS; PR00167; CACCHANNEL.
DR SMART; SM00384; AT_hook_1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Disease mutation.
FT 1
FT REPEAT
FT 427 671
FT REPEAT
FT 1134 1417
FT REPEAT
FT 1454 1717
FT DOMAIN
FT 53 71
FT DOMAIN
FT 72 90
FT TRANSMEM
FT 91 108
FT DOMAIN
FT 109 120
FT TRANSMEM
FT 121 136
FT DOMAIN
FT 137 144
FT TRANSMEM
FT 145 163
FT DOMAIN
FT 164 182
FT TRANSMEM
FT 183 202
FT DOMAIN
FT 203 289
FT TRANSMEM
FT 290 314
FT DOMAIN
FT 315 441
FT TRANSMEM
FT 442 461
FT DOMAIN
FT 462 475
FT TRANSMEM
FT 476 495
FT DOMAIN
FT 496 503
FT TRANSMEM
FT 504 522
FT DOMAIN
FT 523 532
FT TRANSMEM
FT 533 551
FT DOMAIN
FT 552 570
FT TRANSMEM
FT 571 590
FT DOMAIN
FT 591 643
FT TRANSMEM
FT 644 668
FT DOMAIN
FT 669 1142
FT TRANSMEM
FT 1143 1166
FT DOMAIN
FT 1167 1183
FT TRANSMEM
FT 1184 1203
FT DOMAIN
FT 1204 1210
FT TRANSMEM
FT 1211 1234
FT DOMAIN
FT 1235 1245
FT TRANSMEM
FT 1246 1263
FT DOMAIN
FT 1264 1282
FT TRANSMEM
FT 1283 1302
FT DOMAIN
FT 1303 1389
FT TRANSMEM
FT 1390 1414
FT DOMAIN
FT 1415 1470
FT TRANSMEM
FT 1471 1489
FT DOMAIN
FT 1490 1503
FT TRANSMEM
FT 1504 1525
FT TRANSMEM
FT 1526 1532
FT DOMAIN
FT 1533 1552
FT TRANSMEM
FT 1553 1559
FT DOMAIN
FT 1560 1578
FT TRANSMEM
FT 1579 1597
FT DOMAIN
FT 1598 1617
FT TRANSMEM
FT 1618 1689
FT DOMAIN
FT 1690 1715
FT TRANSMEM
FT 1716 2164
FT DOMAIN
FT 681 686
FT DOMAIN
FT 1107 1110
FT TRANSMEM
FT 2114 2123
FT DOMAIN
FT 337 354
FT SITE
FT 272 272
FT SITE
FT 622 622
FT SITE
FT 1363 1363

II.
II.
IV.
S1 OF REPEAT I (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S2 OF REPEAT I (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S3 OF REPEAT I (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S4 OF REPEAT I (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S5 OF REPEAT I (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S6 OF REPEAT I (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S1 OF REPEAT II (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S2 OF REPEAT II (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S3 OF REPEAT II (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S4 OF REPEAT II (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S5 OF REPEAT II (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S6 OF REPEAT II (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S1 OF REPEAT III (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S2 OF REPEAT III (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S3 OF REPEAT III (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S4 OF REPEAT III (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S5 OF REPEAT III (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S6 OF REPEAT III (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S1 OF REPEAT IV (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S2 OF REPEAT IV (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S3 OF REPEAT IV (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S4 OF REPEAT IV (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
S5 OF REPEAT IV (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
S6 OF REPEAT IV (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
POLY-GLU.
POLY-GLU.
BINDING TO THE BETA SUBUNIT (BY
CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).
CALCIUM ION SELECTIVITY AND PERMEABILITY
(BY SIMILARITY).


```

CC SITE 1659 1659 CALCIUM ION SELECTIVITY AND PERMEABILITY
CC (BY SIMILARITY).
CC MOD RRS 1725 1725 PHOSPHORYLATION (BY PKA) (POTENTIAL).
CC CA BIND 1743 1754 BY SIMILARITY.
CC CARBOHYD 237 1237 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1559 1559 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC VARIANT 601 601 P -> L (IN TG).
CC SEQUENCE 2164 AA; 246978 MM; E9A80AC5B4758B8C CRC64;

Query Match 72.7%; Score 32; DB 1; Length 2164;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ISEMFLOIT 9
Db 485 MSEMFIKOT 493

RESULT 6
CCAA_RAT STANDARD; PRT; 2212 AA.
AC P54282; 001541;
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Voltage-dependent P/Q-type calcium channel alpha-1a subunit (Calcium
DE channel 1) (BI) (RAT brain class A) (RBA-1).
GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Brain;
RX MEDLINE=9128516; PubMed=1648226;
RA Starr T.V.B., Pryetay W., Shutch T.P.;
RT "Primary structure of a calcium channel that is highly expressed in
RT the rat cerebellum."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5621-5625 (1991).
RN [2]
RP SEQUENCE OF 1435-1667 FROM N.A. (CLONE RBA-65).
RX MEDLINE=90239020; PubMed=1692134;
RA Shutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.;
RT "Rat brain expresses a heterogeneous family of calcium channels."
RL Proc. Natl. Acad. Sci. U.S.A. 87:3391-3395 (1990).
RN [3]
RP SEQUENCE OF 1354-1659 FROM N.A. (ISOFORM RKC8).
RC TISSUE-Kidney;
RX MEDLINE=9306265; PubMed=1279681;
RA Yu A.S.L., Hebert S.C., Brenner B.M., Lytton J.;
RT "Molecular characterization and nephron distribution of a family of
RT transcripts encoding the pore-forming subunit of Ca2+ channels in the
RT kidney."
RL Proc. Natl. Acad. Sci. U.S.A. 89:10494-10498 (1992).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPIRIDINES (DHP), AND OMEGA-CONOTOXIN-GYIA (OMEGA-CTX-
CC GYIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA

```

```

CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; PURKINE CELLS CONTAIN
CC PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
CC CURRENT IN CEREBELLAR GRANULE CELLS. ALSO FOUND IN HEART, IN
CC KIDNEY DISTAL CONVOLUTED TUBULE (DCT), AND IN PITUITARY.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M64373; AAA40806.1; -.
CC EMBL; M99222; AAA40896.1; -.
CC InterPro: IPR000637; AT hook.
CC InterPro: IPR001682; Ca/Na pore.
CC InterPro: IPR002077; Ca channel.
CC InterPro: IPR002111; Cat channel.
CC InterPro: IPR000636; M+channel_nlg.
CC Pfam; PF00520; ion trans; 4.
CC PRINTS; PR00167; CACCHANNEL.
CC SMART; SM00384; AT hook; 1.
CC -1- Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
CC Calcium channel; Glycoprotein; Repeat; Multigene family;
CC KM Calcium-binding; Phosphorylation; Alternative splicing.
CC KM REPEAT 87 365
CC FT REPEAT 475 719
CC FT REPEAT 1182 1465
CC FT REPEAT 1502 1765
CC FT DOMAIN 1 100
CC FT TRANSSEM 101 119
CC FT DOMAIN 120 138
CC FT TRANSSEM 139 156
CC FT DOMAIN 157 168
CC FT TRANSSEM 169 184
CC FT DOMAIN 185 192
CC FT TRANSSEM 193 211
CC FT DOMAIN 212 230
CC FT TRANSSEM 231 250
CC FT DOMAIN 251 337
CC FT TRANSSEM 338 362
CC FT DOMAIN 363 489
CC FT TRANSSEM 490 509
CC FT DOMAIN 510 523
CC FT TRANSSEM 524 543
CC FT DOMAIN 544 551
CC FT TRANSSEM 552 570
CC FT DOMAIN 571 580
CC FT TRANSSEM 581 599
CC FT DOMAIN 600 618
CC FT TRANSSEM 619 638
CC FT DOMAIN 639 691
CC FT TRANSSEM 692 716
CC FT DOMAIN 717 1190
CC FT TRANSSEM 1191 1214
CC FT DOMAIN 1215 1231
CC FT TRANSSEM 1232 1251
CC FT DOMAIN 1252 1258
CC FT TRANSSEM 1259 1283
CC FT DOMAIN 1283 1293
CC FT TRANSSEM 1294 1311

```


FT DOMAIN 186 190 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 191 209 S4 OF REPEAT I (POTENTIAL).
 FT DOMAIN 210 228 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 229 248 S5 OF REPEAT I (POTENTIAL).
 FT DOMAIN 249 335 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 336 360 S6 OF REPEAT I (POTENTIAL).
 FT DOMAIN 361 487 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 488 506 S1 OF REPEAT II (POTENTIAL).
 FT DOMAIN 507 521 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 522 541 S2 OF REPEAT II (POTENTIAL).
 FT DOMAIN 542 549 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 550 568 S3 OF REPEAT II (POTENTIAL).
 FT DOMAIN 569 578 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 579 597 S4 OF REPEAT II (POTENTIAL).
 FT DOMAIN 598 616 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 617 636 S5 OF REPEAT II (POTENTIAL).
 FT DOMAIN 637 689 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 690 714 S6 OF REPEAT II (POTENTIAL).
 FT DOMAIN 715 1253 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1254 1272 S1 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1273 1288 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1289 1308 S2 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1309 1320 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1321 1339 S3 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1340 1350 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1351 1369 S4 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1370 1388 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1389 1408 S5 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1409 1495 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1496 1520 S6 OF REPEAT III (POTENTIAL).
 FT DOMAIN 1521 1575 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1576 1604 S1 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1605 1609 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1610 1629 S2 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1630 1637 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1638 1656 S3 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1657 1665 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1666 1684 S4 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1685 1703 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1704 1723 S5 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1724 1795 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1796 1820 S6 OF REPEAT IV (POTENTIAL).
 FT DOMAIN 1821 2424 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 13 18 POLY-GLY.
 FT DOMAIN 727 732 POLY-GLY.
 FT TRANSMEM 1004 1010 POLY-ARG.
 FT DOMAIN 1012 1017 POLY-HIS.
 FT TRANSMEM 2219 2227 POLY-ARG.
 FT DOMAIN 2242 2246 POLY-ARG.
 FT TRANSMEM 2288 2297 POLY-ARG.
 FT DOMAIN 2298 2301 POLY-GLY.
 FT TRANSMEM 2372 2377 POLY-PRO.
 FT DOMAIN 2411 2416 POLY-GLY.
 FT TRANSMEM 383 400 BINDING TO THE BETA SUBUNIT.
 FT DOMAIN 318 318 CALCULUM ION SELECTIVITY AND PERMEABILITY
 FT TRANSMEM 668 668 (BY SIMILARITY).
 FT DOMAIN 1469 1469 CALCULUM ION SELECTIVITY AND PERMEABILITY
 FT TRANSMEM 1765 1765 (BY SIMILARITY).
 FT DOMAIN 1831 1831 MOD RES 1831 1831 PHOSPHORYLATION (BY PRA) (POTENTIAL).
 FT CA BIND 1849 1860 BY SIMILARITY.
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1665 1665 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TRANSMEM 772 1051 MISSING (IN ISOFORM CBP103).
 FT VARSPLIC 772 1120 MISSING (IN ISOFORM CBP103).
 FT VARSPLIC 1857 1884 LYSDRYANLRMPPLKGCNCPAYAY -> HYKMYSL
 FT VARSPLIC 2230 2273 RYSPPLGCKCKCHRYAC (IN ISOFORM
 FT VARSPLIC 2230 2273 CBP101).
 FT VARSPLIC 2230 2273 RQPGVSPGVARRRRGPAVAVRPARAPALAHARARAP
 FT VARSPLIC 2230 2273 ARL -> PAADKERYGPODRPDHGRARARADWRMSPS

FT VARSPLIC 2274 2424 EGRHTTHRO (IN ISOFORM BI-1).
 FT TRANSMEM 419 419 MISSING (IN ISOFORM BI-1).
 FT VARIANT 877 877 MISSING (IN ISOFORM CBP315).
 FT VARIANT 1104 1104 A -> T (IN ISOFORM CBS).
 FT MUTAGEN 386 386 S -> N (IN ISOFORM CBS).
 FT MUTAGEN 389 389 E-S: REDUCED BETA-SUBUNIT INTERACTION.
 FT MUTAGEN 392 392 L-S: REDUCED BETA-SUBUNIT INTERACTION.
 FT MUTAGEN 400 400 Y-S: REDUCED BETA-SUBUNIT INTERACTION.
 FT MUTAGEN 400 400 E-S: NO EFFECT ON BETA-SUBUNIT
 FT MUTAGEN 400 400 INTERACTION.
 FT MUTAGEN 400 400 INTERACTION.
 SQ SEQUENCE 2424 AA; 273228 MW; F7CC4D0AB4B45604 CRC64;
 Query Match 72.7%; Score 32; DB 1; Length 2424;
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 ISEMPLOY 9
 Db 531 MSEMFINOV 539
 RESULT 8
 CCAA HUMAN STANDARD; PRT; 2505 AA.
 ID CCAA HUMAN
 AC 00055; Q92690; Q16290; Q99790; Q99791; Q99792; Q99793; P78510;
 AC P78511;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
 DE channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel
 DE I) (BI).
 GN CACNA1A OR CACNA1A4 OR CACNA4 OR CACNA3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS BI-1-GCAG/1A-1 AND BI-1/1A-2).
 RC TISSUE=Neuron;
 RX MEDLINE=99158614; PubMed=10049321;
 RA Hans M., Urrutia A., Deal C., Brust P.F., Strydom K., Ellis S.B.,
 RA Harpold M.W., Johnson E.C., Williams M.E.;
 RT "Structural elements in domain IV that influence biophysical and
 RT pharmacological properties of human alpha1A-containing high-voltage-
 RT activated calcium channels.";
 RL Biophys. J. 76:1384-1400(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM BI-1(VI)), AND VARIANTS PHM.
 RC TISSUE=Cerebellum;
 RX MEDLINE=97053792; PubMed=8898206;
 RA Ophoff R.A., Terwindt G.M., Vergouwe M.N., van Eijk R., Oefner P.J.,
 RA Hoffman S.M.G., Lamerdin J.E., Mohrenwieser H.W., Bulman D.E.,
 RA Ferrari M.D., Haan J., Lindhout D., van Ommen G.-J.B., Hofker M.H.,
 RA Ferrari M.D., Frants R.R.;
 RT "Familial hemiplegic migraine and episodic ataxia type-2 are caused by
 RT mutations in the Ca2+ channel gene CACNA1A.";
 RL Cell 87:543-552(1996).
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=97141920; PubMed=8988170;
 RA Zhuchenko O., Bailey J., Bonnen P.E., Ashizawa T., Stockton D.W.,
 RA Amos C., Dobyns W.B., Subramony S.H., Zoghbi H.Y., Lee C.C.;
 RT "Autosomal dominant cerebellar ataxia (SCA6) associated with small
 RT polyglutamine expansions in the alpha 1A-voltage-dependent calcium
 RT channel.";
 RL Nat. Genet. 15:62-69(1997).
 RP SEQUENCE OF 1233-1651 FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stiellwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,

RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
 RA Kobayashi A., Olsen A.S., Carraro A.V.,
 RA Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1693-1807 FROM N.A.
 RC TISSUE=Lung carcinoma;
 RA MEDLINE=9512349; PubMed=782133;
 RA Barry E.L.R., Viglione M.P., Kim Y.I., Froehner S.C.;
 RT "Expression and antibody inhibition of P-type calcium channels in
 RT human small-cell lung carcinoma cells.";
 RL J. Neurosci. 15:274-283(1995).
 RN [6]
 RP SEQUENCE OF 2038-2258 FROM N.A.
 RC TISSUE=Frontal cortex;
 RA MEDLINE=96102310; PubMed=8525433;
 RA Margolis R.L., Breschel T.S., Li S.H., Kidwai A.S., Antonarakis S.E.,
 RA McInnis M.G., Rose C.A., Li S.H., Kidwai A.S., Antonarakis S.E.,
 RT "Characterization of cDNA clones containing CCA trinucleotide repeats
 RT derived from human brain.";
 RL Somat. Cell Mol. Genet. 21:279-284(1995).
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
 CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
 CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
 CC GIVES RISE TO P AND/OR O-TYPE CALCIUM CURRENTS. P/O-TYPE CALCIUM
 CC CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
 CC ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AAROTOXIN-
 CC IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
 CC DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
 CC GVIA).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; B1-1/1A-2, B1-1(V1),
 CC B1-1/GGAG/1A-1 (SHOWN HERE), B1-1(V1)-GGGAG, B1-1(V2), B1-1(V2)-
 CC GGAG AND B1-1(V2,V3); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN CEREBELLUM,
 CC CEREBRAL CORTEX, THALAMUS AND HYPOTHALAMUS. NO EXPRESSION IN
 CC HEART, KIDNEY, LIVER OR MUSCLE. PURKINJE CELLS CONTAIN
 CC PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
 CC CURRENT IN CEREBELLAR GRANULE CELLS.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- POLYMORPHISM: THE POLY-GLN REGION OF CANCA1A IS POLYMORPHIC: 6 TO
 CC 17 REPEATS IN THE NORMAL POPULATION, EXPANDED TO ABOUT 21 TO 30
 CC REPEATS IN SPINOCEREBELLAR ATAXIA 6 (SC6) PATIENTS. THERE SEEMS
 CC TO BE A CORRELATION BETWEEN THE REPEAT NUMBER AND EARLIER ONSET OF
 CC THE DISORDER.
 CC -1- DISEASE: EXPANSION OF A CAG REPEAT IN THE CODING REGION OF CANCA1A
 CC IS THE CAUSE OF ONE FORM OF SPINOCEREBELLAR ATAXIA 6 (SC6). AN
 CC AUTOSOMAL DOMINANT DISORDER CHARACTERIZED BY SLOWLY PROGRESSIVE
 CC CEREBELLAR ATAXIA OF THE LIMBS AND GAIT, DYSARTHRIA, NYSTAGMUS
 CC AND MILD VIBRATORY AND PROPRIOCEPTIVE SENSORY LOSS. THESE SYMPTOMS
 CC ARE PROBABLY EXPLAINED BY SEVERE LOSS OF CEREBELLAR PURKINJE
 CC CELLS.
 CC -1- DISEASE: DEFECTS IN CANCA1A ARE THE CAUSE OF FAMILIAL HEMIPLEGIC
 CC MIGRAINE (FHM), ALSO KNOWN AS MIGRAINE FAMILIAL HEMIPLEGIC 1
 CC (MHP1), EPISODIC ATAXIA TYPE 2 (EA-2), ACETAZOLAMIDE-RESPONSIVE
 CC HEREDITARY PAROXYSMAL CEREBELLAR ATAXIA (APCA), AND POSSIBLY OTHER
 CC COMMON TYPES OF MIGRAINES. FHM, A RARE AUTOSOMAL DOMINANT SUBTYPE
 CC OF MIGRAINE WITH AURA, IS ASSOCIATED WITH ICTAL HEMIPARESIS AND,

CC IN SOME FAMILIES, PROGRESSIVE CEREBELLAR ATROPHY. EA-2 IS ANOTHER
 CC AUTOSOMAL DOMINANT PAROXYSMAL CEREBELLAR DISEASE, CHARACTERIZED BY
 CC ACTOCOLAMIDE-RESPONSIVE ATTACKS OF CEREBELLAR ATAXIA AND
 CC MIGRAINE-LIKE SYMPTOMS, INTERICTAL NYSTAGMUS, AND CEREBELLAR
 CC ATROPHY.
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF004884; AAB61613.1; -;
 CC EMBL: AF004883; AAB61612.1; -;
 CC EMBL: X99897; CA68172.1; -;
 CC EMBL: Z80115; -; NOT_ANNOTATED_CDS.
 CC EMBL: Z80115; -; NOT_ANNOTATED_CDS.
 CC EMBL: U79666; AAB64179.1; -;
 CC EMBL: U79663; AAB49674.1; ALT_INIT.
 CC EMBL: U79664; AAB49675.1; ALT_INIT.
 CC EMBL: U79665; AAB49676.1; ALT_INIT.
 CC EMBL: U79667; AAB49677.1; ALT_INIT.
 CC EMBL: U79668; AAB49678.1; ALT_INIT.
 CC EMBL: AC005305; AAC26839.1; -;
 CC EMBL: S76537; AAB33068.1; -;
 CC EMBL: U06702; -; NOT_ANNOTATED_CDS.
 CC Gene: HGNC:1388; CANCA1A.
 CC MIM: 601011; -;
 CC MIM: 183086; -;
 CC MIM: 141500; -;
 CC MIM: 108500; -;
 CC InterPro: IPR001682; Ca/Na_pore.
 CC InterPro: IPR002077; Ca_channel.
 CC InterPro: IPR002111; Cat_channel_TrypL.
 CC InterPro: IPR000636; M-channel_nlg.
 CC Pfam: PF00520; Ion_trans_4.
 CC PRINTS: PR00167; CACCHANNEL.
 CC KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 CC KW Calcium channel; Glycoprotein; Repeat; Multigene family; Polymorphism;
 CC KW Calcium-binding; Phosphorylation; Alternative splicing; Polymorphism;
 CC FT Disease mutation; Triplet repeat expansion.
 CC FT REPEAT 85 363 1.
 CC FT REPEAT 473 717 II.
 CC FT REPEAT 1231 1514 III.
 CC FT REPEAT 1551 1814 IV.
 CC FT DOMAIN 1 98 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 99 117 S1 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 118 135 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 136 155 S2 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 156 167 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 168 185 S3 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 186 190 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 191 209 S4 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 210 228 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 229 248 S5 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 249 335 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 336 360 S6 OF REPEAT I (POTENTIAL).
 CC FT DOMAIN 361 487 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 488 506 S1 OF REPEAT II (POTENTIAL).
 CC FT DOMAIN 507 521 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 522 541 S2 OF REPEAT II (POTENTIAL).
 CC FT DOMAIN 542 549 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 550 568 S3 OF REPEAT II (POTENTIAL).
 CC FT DOMAIN 569 578 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 579 597 S4 OF REPEAT II (POTENTIAL).
 CC FT DOMAIN 598 616 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 617 636 S5 OF REPEAT II (POTENTIAL).
 CC FT DOMAIN 637 689 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 690 714 S6 OF REPEAT II (POTENTIAL).

FT DOMAIN 715 1242 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1243 1261 S1 OF REPEAT II (POTENTIAL).
 FT DOMAIN 1262 1277 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1278 1297 S2 OF REPEAT III (POTENTIAL).

Query Match
 Best Local Similarity 72.7%; Score 32; DB 1; Length 2505;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
 Db 531 MSEMFLQY 539

RESULT 9

SODM_PALVU STANDARD; PRT; 144 AA.
 AC P28765.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Superoxide dismutase [Mn], mitochondrial (EC 1.15.1.1) (Fragment).
 OS *Palinurus vulgaris* (European spiny lobster).
 OC *Palinurus vulgaris* (Arthropoda: Mandibulata); Pancrustacea; Crustacea;
 OC Eucarystacea; Eumalacostraca; Eucarida; Decapoda; Placemata;
 OC Palinura; Palinuroidea; Palinuridae; Palinurus.
 OC NCBI_TaxID=6733;

RP SEQUENCE FROM N.A.
 RX MEDLINE=92211732; PubMed=1556751;
 RA Smith M.W., Doolittle R.F.;
 RT "A comparison of evolutionary rates of the two major kinds of
 RT superoxide dismutase."
 RL J. Mol. Evol. 34:175-184 (1992).
 CC -1- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COFACTOR: Manganese (By similarity).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: X64063; CA45419.1; -
 CC PIR: S23659; S23659.
 CC DR HSPB; P04179; IAP6.
 CC DR InterPro: IPR001189; SODismutase.
 CC DR Pfam: PF00081; sodef; 1.
 CC DR Pfam: PF02777; sodef; 1.
 CC DR Prodom: PD000475; SODismutase; 1.
 CC DR PROSITE: PS00088; SOD_MN; PARTIAL.
 CC KW Oxidoreductase; Manganese; Mitochondrion.
 CC FT NON_TER 1
 CC FT METAL 10
 CC FT METAL 58
 CC FT METAL 143
 CC FT NON_TER 144
 CC FT SEQUENCE 144 AA; 15505 MW; F180CE523653686 CRC64;

Query Match
 Best Local Similarity 70.5%; Score 31; DB 1; Length 144;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
 Db 531 MSEMFLQY 539

Db 2 ISEMFLQY 10

RESULT 10
 VGLM_HSVSA STANDARD; PRT; 366 AA.
 AC 001017;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Glycoprotein M.
 GN 39.
 OS Herpesvirus saimiri (strain 11).
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OC NCBI_TaxID=10383;
 OC [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=9233668; PubMed=1321287;
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.W.;
 RT "Primary structure of the herpesvirus saimiri genome."
 RL J. Virol. 66:5047-5058 (1992).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN M.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: X64346; CA45662.1; -
 CC DR PIR: A36810; OGBHNS.
 CC DR InterPro: IPR000785; Herpes_glycop.
 CC DR Pfam: PF01528; Herpes_glycop; 1.
 CC DR PRINTS: PR00333; HSVINTEGRAMP.
 CC KW Transmembrane; Glycoprotein.
 CC FT TRANSMEM 18
 CC FT TRANSMEM 34
 CC FT TRANSMEM 87
 CC FT TRANSMEM 103
 CC FT TRANSMEM 152
 CC FT TRANSMEM 168
 CC FT TRANSMEM 214
 CC FT TRANSMEM 230
 CC FT TRANSMEM 236
 CC FT TRANSMEM 252
 CC FT TRANSMEM 274
 CC FT TRANSMEM 290
 CC FT TRANSMEM 305
 CC FT TRANSMEM 321
 CC FT DOMAIN 342
 CC FT CANBOHYD 57
 CC FT SEQUENCE 366 AA; 42182 MW; 4EAE0C3639E76463 CRC64;

Query Match
 Best Local Similarity 70.5%; Score 31; DB 1; Length 366;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
 Db 259 MTEMFLQY 267

RESULT 11
 YOB4_MYCTU STANDARD; PRT; 491 AA.
 ID YOB4_MYCTU
 AC O53209;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein Y2484C.
 GN Y2484C OR MT2557 OR KMY008.40C.
 OS Mycobacterium tuberculosis.
 OS Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

Qy 1 ISEMFLQY 9
 Db 259 MTEMFLQY 267

```

OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634220;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Dwyer R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson M.C., Mayhew L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021246; CAA16061.1; -
DR EMBL; AE007092; AAK46861.1; -
DR TIGR; MT2557; -
DR Tuberculist; RV2484C; -
DR InterPro; IPRO04255; UPF0089.
DR Pfam; PF03007; UPF0089; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 491 AA; 52309 MW; C2ACCA458ADF82DD CRC64;
Query Match 70.5%; Score 31; DB 1; Length 491;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 3 EMPLOY 9
Db 155 EMPLOY 161
RESULT 12
YEJA_ECOLI
ID YEJA_ECOLI STANDARD; PRT; 604 AA.
AC P33913; P76447;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yeja precursor.
GN YEJA OR B2177.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

```

```

RA Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.O.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 193-604 FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richierich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE OF 193-604 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alpa H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Mocomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horikuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- SIMILARITY: SOME, TO H. INFLUENZAE HBPA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000307; AAC75238.1; ALT_INIT.
DR EMBL; U00008; AAA16375.1; -
DR EMBL; D90849; BAA15985.1; ALT_INIT.
DR Ecogen; EG12037; yeja.
DR InterPro; IPRO00914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 2.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 604 HYPOTHETICAL PROTEIN YEJA.
FT CONFLICT 193 195 YMK -> IRP (IN REF. 2).
FT CONFLICT 415 415 W -> G (IN REF. 2).
SQ SEQUENCE 604 AA; 69669 MW; E3D45FAF94F16B47 CRC64;
Query Match 70.5%; Score 31; DB 1; Length 604;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 SEMELOY 9
Db 381 SEVLOY 388
RESULT 13
APCE_SYNY4
ID APCE_SYNY4 STANDARD; PRT; 896 AA.
AC 002907;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Phycobillosome 100.5 kDa core-membrane linker polypeptide (L-CM 100.5)
DE (Anchor polypeptide).
OS Synecocystis sp. (strain PCC 6714).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93222481; PubMed=8467079;
RA Dimagno L.M., Haselkorn R.;
RT "Isolation and characterization of the genes encoding allophycocyanin

```

RT subunits and two linker proteins from *Synechocystis* 6714.";
RL Plant Mol. Biol. 21:835-846(1993).
CC -1- FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL
CC ENERGY ACCEPTOR (BY ITS PHYCOBILIN-LIKE DOMAIN) AND AS A LINKER
CC POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE
CC PHYCOBILISOME CORE ARCHITECTURE.
CC -1- SUBUNIT: PHYCOBILISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO
CC CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY
CC COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS AND OF MINOR
CC COMPONENTS.
CC -1- SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILISOME PERPENDICULARLY
CC TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC -1- SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL
CC REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDES.
CC -1- SIMILARITY: THE PHYCOBILIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILINS
CC FROM VARIOUS SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 102309; AAA69685.1; -;
DR HSSP; P00318; 1B33.
DR InterPro; IPR001297; PBS_linker_poly.
DR InterPro; IPR001659; Phycobillisme.
DR Pfam; PF00427; PBS_linker_poly; 3.
DR Pfam; PF00502; Phycobillisme; 1.
DR ProDom; PD000340; Phycobillisme; 2.
DR Phycobillisme; Electron transport; Photosynthesis; Repeat.
FT DOMAIN 233 280 ARM 1 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 281 405 1.
FT DOMAIN 406 541 ARM 2 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 542 664 11.
FT DOMAIN 665 737 ARM 3 (SPACING SEQUENCE) (BY SIMILARITY).
FT REPEAT 738 863 111.
SQ SEQUENCE 896 AA; 100460 MW; BDID14C2923E32A3 CRC64;
Query Match 70.5%; Score 31; DB 1; Length 896;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 SEMPLQTY 9
Db 585 SELFLKLY 592
RESULT 14
PR16_YEAST
ID PR16_YEAST STANDARD; PRT; 1071 AA.
AC P15936;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pre-mRNA splicing factor RNA helicase PRP16.
GN PRP16 OR YKR086W OR YKR406.
OS *Saccharomyces cerevisiae* (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90182661; PubMed=2138057;
RA Burgess S., Couto J.R., Guthrie C.;
RT "A putative ATP binding protein influences the fidelity of
RT branchpoint recognition in yeast splicing.";
RL Cell 60:705-717(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262327; PubMed=8203164;

RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
RA Remacha M.A., Revuelta J.P., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of *Saccharomyces*
RT *cerevisiae* chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
RT and six new open reading frames.";
RL Yeast 10:231-245(1994).
RN [3]
RP FUNCTION.
RX MEDLINE=91125476; PubMed=1825134;
RA Scher B., Guthrie C.;
RT "PRP16 is an RNA-dependent ATPase that interacts transiently with the
RT spliceosome.";
RL Nature 349:494-499(1991).
CC -1- FUNCTION: INFLUENCES THE FIDELITY OF BRANCHPOINT RECOGNITION IN
CC YEAST SPLICING. THIS IS RNA-DEPENDENT ATPASE WHICH IS ESSENTIAL
CC FOR VIABILITY. IT MAY MEDIATE ONE OF THE MANY ATP-REQUIRING STEPS
CC OF SPliceosome ASSEMBLY AND THAT ACCURACY OF BRANCHPOINT
CC RECOGNITION MAY BE COUPLED TO ATP BINDING AND/OR HYDROLYSIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. BINDS TO THE SPLICOSOME.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
CC SUBFAMILY. PRP16 ORTHOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31524; AAA34911.1; -;
DR EMBL; 227116; CAA81637.1; -;
DR EMBL; 228311; CAA82165.1; -;
DR PIR; S38164; S38164.
DR PIR; S39127; S39127.
DR SGD; S0001794; PRP16.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SMO0487; DEXDC; 1.
DR SMART; SMO0490; HELIC; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; 1.
DR mRNA processing; mRNA splicing; Helicase; ATP-binding;
KW Nuclear protein.
FT NP_BIND 373 380 ATP (BY SIMILARITY).
FT SITE 473 476 DEAD BOX.
FT MUTAGEN 386 386 Y->D: SUPPRESSOR PHENOTYPE.
FT CONFLICT 698 698 A -> R (IN REF. 1).
SQ SEQUENCE 1071 AA; 121652 MW; EB7D4C4A18F5D4E1 CRC64;
Query Match 70.5%; Score 31; DB 1; Length 1071;
Best Local Similarity 55.6%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 SEMPLQTY 9
Db 583 LOEXFLQY 591
RESULT 15
MSP1_PLAFK
ID MSP1_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merzoitte surface protein 1 precursor (Merzoitte surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS *Plasmodium falciparum* (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stuenkelberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS. SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential1).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1; -.
DR PIR; A25120; SA2OK1.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyporein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1630
FT DOMAIN 67 84
FT TRANSMEM 1614 1630
FT CARBOHYD 97 97
FT CARBOHYD 259 259
FT CARBOHYD 755 755
FT CARBOHYD 774 774
FT CARBOHYD 835 835
FT CARBOHYD 911 911
FT CARBOHYD 955 955
FT CARBOHYD 1049 1049
FT CARBOHYD 1156 1156
FT CARBOHYD 1165 1165
FT CARBOHYD 1436 1436
FT CARBOHYD 1517 1517
SQ SEQUENCE 1630 AA; 187289 MW; ADHDEC3CE0A46322 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 1630;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
:|:|:|:|:|
Db 838 LSQLFMEIY 846

Search completed: May 1, 2003, 23:10:57
Job time : 3.32143 secs

```


Thu May 8 16:14:15 2003

us-09-658-621b-26.rapb

Page 1

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 1, 2003, 23:12:56 ; Search time 0.571429 Seconds

(Without alignments)
1359.015 Million cell updates/sec

Title: US-09-658-621b-26

Perfect score: 44

Sequence: 1 ISEMFLQIY 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328255 seqs, 8628685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodaa/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodaa/2/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodaa/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodaa/2/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodaa/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodaa/2/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodaa/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodaa/2/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodaa/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodaa/2/pubpaa/US09_NEW_PUB.pep.*
11: /cgn2_6/ptodaa/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodaa/2/pubpaa/US10_NEW_PUB.pep.*
13: /cgn2_6/ptodaa/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodaa/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 44 | 100.0 | 321 | 10 | US-09-925-301-861 |
| 2 | 44 | 100.0 | 1255 | 9 | US-09-996-069-10 |
| 3 | 41 | 93.2 | 508 | 9 | US-10-057-136-20 |
| 4 | 31 | 70.5 | 1639 | 9 | US-10-087-464-10 |
| 5 | 30 | 68.2 | 207 | 9 | US-09-963-959-7 |
| 6 | 30 | 68.2 | 272 | 9 | US-10-060-019-27 |
| 7 | 30 | 68.2 | 815 | 9 | US-10-060-019-24 |
| 8 | 29 | 65.9 | 44 | 9 | US-09-972-970-5 |
| 9 | 29 | 65.9 | 69 | 9 | US-10-091-504-1005 |
| 10 | 29 | 65.9 | 69 | 10 | US-09-764-869-1005 |
| 11 | 29 | 65.9 | 485 | 10 | US-09-925-300-1664 |
| 12 | 28 | 63.6 | 111 | 9 | US-10-078-090-209 |
| 13 | 28 | 63.6 | 261 | 9 | US-10-166-087-34 |
| 14 | 28 | 63.6 | 299 | 9 | US-09-510-332-1 |
| 15 | 28 | 63.6 | 299 | 10 | US-09-593-634-35 |
| 16 | 28 | 63.6 | 332 | 9 | US-10-043-487-11 |
| 17 | 28 | 63.6 | 349 | 10 | US-09-804-551B-4 |
| 18 | 28 | 63.6 | 353 | 10 | US-09-841-132-510 |
| 19 | 28 | 63.6 | 410 | 9 | US-09-987-021-3 |

| | | | | | | |
|----|----|------|------|----|--------------------|--------------------|
| 20 | 28 | 63.6 | 410 | 10 | US-09-957-485-3 | Sequence 3, App1 |
| 21 | 28 | 63.6 | 662 | 12 | US-10-036-328A-8 | Sequence 8, App1 |
| 22 | 28 | 63.6 | 714 | 10 | US-09-818-264-2 | Sequence 2, App1 |
| 23 | 28 | 63.6 | 733 | 12 | US-10-036-328A-4 | Sequence 4, App1 |
| 24 | 28 | 63.6 | 829 | 9 | US-10-102-806-602 | Sequence 602, App1 |
| 25 | 28 | 63.6 | 1029 | 9 | US-10-033-245-22 | Sequence 22, App1 |
| 26 | 28 | 63.6 | 1029 | 9 | US-10-033-223-22 | Sequence 22, App1 |
| 27 | 28 | 63.6 | 1029 | 9 | US-10-033-167-22 | Sequence 22, App1 |
| 28 | 28 | 63.6 | 1029 | 9 | US-10-033-244-22 | Sequence 22, App1 |
| 29 | 28 | 63.6 | 1029 | 9 | US-10-033-435-22 | Sequence 22, App1 |
| 30 | 28 | 63.6 | 1029 | 9 | US-10-032-990-22 | Sequence 22, App1 |
| 31 | 28 | 63.6 | 1029 | 9 | US-09-929-769-7 | Sequence 7, App1 |
| 32 | 28 | 63.6 | 1029 | 9 | US-10-032-996-22 | Sequence 22, App1 |
| 33 | 28 | 63.6 | 1029 | 9 | US-10-033-396-22 | Sequence 22, App1 |
| 34 | 28 | 63.6 | 1029 | 12 | US-10-033-246-22 | Sequence 22, App1 |
| 35 | 28 | 63.6 | 1029 | 12 | US-10-033-301-22 | Sequence 22, App1 |
| 36 | 28 | 63.6 | 1029 | 12 | US-10-033-326-22 | Sequence 22, App1 |
| 37 | 28 | 63.6 | 1138 | 12 | US-10-036-328A-6 | Sequence 6, App1 |
| 38 | 28 | 63.6 | 1210 | 12 | US-10-036-328A-2 | Sequence 2, App1 |
| 39 | 27 | 61.4 | 61 | 9 | US-10-092-154-566 | Sequence 566, App |
| 40 | 27 | 61.4 | 61 | 10 | US-09-764-847-566 | Sequence 166, App |
| 41 | 27 | 61.4 | 67 | 10 | US-09-729-674-166 | Sequence 5152, App |
| 42 | 27 | 61.4 | 84 | 9 | US-09-764-891-5152 | Sequence 976, App |
| 43 | 27 | 61.4 | 88 | 9 | US-10-092-154-976 | Sequence 976, App |
| 44 | 27 | 61.4 | 88 | 10 | US-09-764-847-976 | Sequence 1370, App |
| 45 | 27 | 61.4 | 168 | 10 | US-09-925-300-1370 | |

ALIGNMENTS

RESULT 1
US-09-925-301-861
Sequence 861, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P1106
CURRENT APPLICATION NUMBER: US/09/925,301
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 861.
LENGTH: 321
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (18)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Query Match 100.0%; Score 44; DB 10; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ISEMFLQIY 9
Db 139 ISEMFLQIY 147
RESULT 2
US-09-996-069-10
Sequence 10, Application US/09996069
GENERAL INFORMATION:
APPLICANT: Bamdad, Cynthia
APPLICANT: Bamdad, R. Shoshana

```
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBIT
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

Query Match          100.0%; Score 44; DB 9; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 1073 ISEMFLQIY 1081

RESULT 3
US-10-057-136-20
; Sequence 20, Application US/10057336
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20

Query Match          93.2%; Score 41; DB 9; Length 508;
Best Local Similarity 88.9%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 326 ISEMFLQIY 334

RESULT 4
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Aghar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goei, Yikias
; APPLICANT: Li, Auecong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
```

```
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match          70.5%; Score 31; DB 9; Length 1639;
Best Local Similarity 44.4%; Pred. No. 4.4e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 847 ISEMFLQIY 855

RESULT 5
US-09-963-959-7
; Sequence 7, Application US/09963959
; Patent No. US20020165145A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 33521, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FAMILY MEMBER
; FILE REFERENCE: 10448-095001
; CURRENT APPLICATION NUMBER: US/09/963,959
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/235,033
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-09-963-959-7

Query Match          68.2%; Score 30; DB 9; Length 207;
Best Local Similarity 55.6%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ISEMFLQIY 9
Db 96 ISEMFLQIY 104

RESULT 6
US-10-060-019-27
; Sequence 27, Application US/10060019
; Publication No. US20030003564A1
; GENERAL INFORMATION:
; APPLICANT: Tyers, Mike
; APPLICANT: Willems, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/10/060,019
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US/09/177,165
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
```

LENGTH: 272
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-060-019-27

Query Match 68.2%; Score 30; DB 9; Length 272;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SEMFLOQY 9
Db 98 SEFLOQY 105

RESULT 7
US-10-060-019-24
Sequence 24, Application US/10060019
Publication No. US20030003564A1
GENERAL INFORMATION:
APPLICANT: Tyers, Mike
APPLICANT: Williams, Andrew
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
FILE REFERENCE: 11757.10USU1
CURRENT APPLICATION NUMBER: US/10/060,019
PRIOR FILING DATE: 1998-10-22 97,165
PRIOR APPLICATION NUMBER: 60/092,443
PRIOR FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: 60/063,254
PRIOR FILING DATE: 1997-10-24
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 815
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-060-019-24

Query Match 68.2%; Score 30; DB 9; Length 815;
Best Local Similarity 75.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SEMFLOQY 9
Db 106 SEFLOQY 113

RESULT 8
US-09-972-970-5
Sequence 5, Application US/09972970
Patent No. US2002016463A1
GENERAL INFORMATION:
APPLICANT: Shi et al.
TITLE OF INVENTION: TMA5F Receptor Polynucleotides, Polypeptides, and Antibodies
FILE REFERENCE: PT056P1
CURRENT APPLICATION NUMBER: US/09/972,970
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: PCT/US01/11130
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/195,336
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 44
TYPE: PRT
ORGANISM: Homo sapiens
US-09-972-970-5

Query Match 65.9%; Score 29; DB 9; Length 44;
Best Local Similarity 44.4%; Pred. No. 30;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ISEMFLQY 9
Db 25 LSDFLQY 33

RESULT 9
US-10-091-504-1005
Sequence 1005, Application US/10091504
Publication No. US2003005908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1005
LENGTH: 69
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (48)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: misc_feature
LOCATION: (49)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: misc_feature
LOCATION: (50)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-1005

Query Match 65.9%; Score 29; DB 9; Length 69;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SEMFLOQY 9
Db 16 SDYFLOQY 23

RESULT 10
US-09-764-869-1005
Sequence 1005, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1005
LENGTH: 69
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (48)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (49)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (50)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-1005

Thu May 8 16:14:15 2003

us-09-658-621b-26.raph

Page 4

| | | | | |
|-------------------------|--------|---------------|---------------|------------|
| Query Match | 65.9%; | Score 29; | DB 10; | Length 69; |
| Best Local Similarity | 62.5%; | Pred. No. 48; | | |
| Matches 5; Conservative | | 2; | Mismatches 1; | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|----|-----------|----|
| QY | 2 | SEMFLOQY | 9 |
| | | : : | |
| Db | 16 | SDYFLQLY | 23 |

```

RESULT 11
US-09-925-300-1664
? Sequence 1664, Application US/09925300
? Patent No. US2002015168A1
? GENERAL INFORMATION:
? APPLICANT: Craig Rosen,
? APPLICANT: Steve Ruben,
? TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
? FILE REFERENCE: P101
? CURRENT APPLICATION NUMBER: US/09/925,300
? CURRENT FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: PCT/US00/05988
? PRIOR FILING DATE: 2000-03-08
? PRIOR APPLICATION NUMBER: 60/124,270
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 1890
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO: 1664
? LENGTH: 485
? TYPE: prt
? ORGANISM: Homo sapiens
? US-09-925-300-1664

```

| | | | | |
|-----------------------|--------|--------------------|--------|---------------|
| Query Match | 65.9% | Score 29; | DB 10; | Length 485; |
| Best Local Similarity | 55.6%; | Pred. No. 3.3e+02; | | |
| Matches | 5; | Conservative | 2; | Mismatches 2; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

```
QY      1 ISEMFLOIY 9
Db      451 IKQMFQOMY 455
```

```

RESULT 12
US-10-078-090-209
: Sequence 209, Application US/10078090
: Publication No., US20030044815A1
: GENERAL INFORMATION:
: APPLICANT: Salceda, Susana
: APPLICANT: Macina, Roberto
: APPLICANT: Hu, Ping
: APPLICANT: Reclipon, Herve
: APPLICANT: Karra, Kalpana
: APPLICANT: Cafferey, Robert
: APPLICANT: Sun, Yongming
: APPLICANT: Liu, Chenghua
: TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Proteins
: FILE REFERENCE: DEX-0312
: CURRENT APPLICATION NUMBER: US/10/078, 090
: CURRENT FILING DATE: 2002-02-14
: PRIOR APPLICATION NUMBER: 60/268, 999
: PRIOR FILING DATE: 2001-02-15
: NUMBER OF SEQ ID NOS: 210
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 209
: LENGTH: 111
: TYPE: PRT
: ORGANISM: Homo sapien
: US-10-078-090-209

```

| | | | | |
|-----------------------|--------|--------------------|-------|---------------------------------|
| Query Match | 63.6%; | Score 28; | DB 9; | Length 11; |
| Best Local Similarity | 66.7%; | Pred. No. 1.2e+02; | | |
| Matches | 6; | Conservative | 0; | Mismatches 3; Indels 0; Gaps 0; |

| | QY | 1 | ISEMFLQIY | 9 |
|----|----|-----------|-----------|---|
| Db | 20 | ISSTFLSIY | 28 | |

```

RESULT 13
US-10-166-087-34
: Sequence 34, Application US/10166087
: Publication No. US20030077767A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Ecopia Biosciences Inc.
: APPLICANT: Farnel, Chris
: APPLICANT: Staiff, Aliredo
: APPLICANT: Zaopoulos, Emmanuel
: TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin
: FILE REFERENCE: 3014-2US
:
: CURRENT APPLICATION NUMBER: US/10/166,087
: CURRENT FILING DATE: 2002-06-11
: NUMBER OF SEQ ID NOS: 51
:
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 34
:
: LENGTH: 261
:
: TYPE: PRT
:
: ORGANISM: Streptomyces rufinus subspecies thermotolerans
:
: US-10-166-087-34

```

| | | | | |
|-----------------------|-------|--------------|---------|--------------|
| Query Match | 63.6% | Score 28 | DB 9 | Length 261 |
| Best Local Similarity | 55.6% | Pred. NO. | 2.9e+02 | |
| Matches | 5 | Conservative | 3 | Mismatches 1 |
| | | | | Indels 0 |
| | | | | Gaps 0 |

```
QY      1 ISEMFLOIY 9
Db      181 IEEVWLEIY 189
```

```

RESULT 14
US-09-510-332-1
; Sequence 1, Application US/09510332
; Publication No. US20030022278A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Meisler, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: T2R, a No. US20030022278A1e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098010US
; CURRENT APPLICATION NUMBER: US/09/510.332
; PRIOR APPLICATION NUMBER: 2000-02-22
; PRIOR APPLICATION NUMBER: US 09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human T2R01 (hGR01)
; US-09-510-332-1

```

| | | | | |
|-----------------------|--------|--------------------|-------|---------------|
| Query Match | 63.6%; | Score 28; | DB 9; | Length 299; |
| Best Local Similarity | 44.4%; | Pred. No. 3.3e+02; | | |
| Matches | 4; | Conservative | 4; | Mismatches 1; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

```
QY      1 ISEMFLOIY 9
Db      53 VSRIFLOLF 61
```

RESULT 15
US-09-393-634-35

```

; Sequence 35, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors
; FILE REFERENCE: 023078-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR01, SF01
US-09-393-634-35

Query Match 63.6%; Score 28; DB 10; Length 299;
Best Local Similarity 44.4%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEWFLQIY 9
: * : | | | :
Db 53 VSRIFLQLF 61

Search completed: May 1, 2003, 23:21:44
Job time : 1.57143 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: May 1, 2003, 23:21:06 ; Search time 15 Seconds
(without alignments)
57.681 Million cell updates/sec

Title: US-09-658-621B-26

Perfect score: 44

Sequence: 1 ISEMFLQIY 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 789

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|------------|---------------------|
| 1 | 17 | 38.6 | 9 2 PT0315 | Ig heavy chain CRD |
| 2 | 15 | 34.1 | 6 2 A27696 | contraction-inhibi |
| 3 | 15 | 34.1 | 7 2 S08606 | hypothetical prote |
| 4 | 15 | 34.1 | 9 2 PT0225 | seminal vesicle pr |
| 5 | 13 | 29.5 | 9 2 PT0225 | Ig heavy chain CDR |
| 6 | 13 | 29.5 | 2 154906 | bone gla protein - |
| 7 | 12 | 27.3 | 4 2 I54357 | schwannomn - mous |
| 8 | 12 | 27.3 | 6 2 B27696 | contraction-inhibi |
| 9 | 12 | 27.3 | 7 2 B39127 | phosphotransferase |
| 10 | 12 | 27.3 | 8 2 A41117 | acetylcholinestera |
| 11 | 12 | 27.3 | 8 2 T13818 | cytochrome oxidase |
| 12 | 12 | 27.3 | 9 2 D58503 | translucine elonga |
| 13 | 12 | 27.3 | 9 2 B41978 | calliFERamide 2 - |
| 14 | 12 | 27.3 | 9 2 C41978 | neuropeptide Antho |
| 15 | 11 | 25.0 | 4 2 A35779 | actin I - malaria |
| 16 | 11 | 25.0 | 5 2 B45525 | phytosulfokine alp |
| 17 | 11 | 25.0 | 5 3 JT0870 | RNA-directed DNA p |
| 18 | 11 | 25.0 | 7 1 NYPG7 | hypothetical prote |
| 19 | 11 | 25.0 | 7 2 B35890 | pulvinase (EC 3. |
| 20 | 11 | 25.0 | 7 2 S09652 | cardiacskeletalory |
| 21 | 11 | 25.0 | 8 2 S66466 | penalbumin - Adeli |
| 22 | 11 | 25.0 | 8 2 A61467 | neural proteinase |
| 23 | 11 | 25.0 | 8 2 S70334 | endosperm protein, |
| 24 | 11 | 25.0 | 8 2 S69478 | globulin IV, alpha |
| 25 | 10 | 22.7 | 4 2 B37325 | pap fibmbrial regul |
| 26 | 10 | 22.7 | 5 2 S70615 | endo-1,4-beta-xyla |
| 27 | 10 | 22.7 | 5 2 B61445 | len-enkephalin - b |
| 28 | 10 | 22.7 | 5 2 B61445 | |

| | | | | |
|----|----|------|------------|--------------------|
| 30 | 10 | 22.7 | 6 2 JN0861 | peptidyl-dipectida |
| 31 | 10 | 22.7 | 6 2 B34835 | dna protein - pse |
| 32 | 10 | 22.7 | 6 2 I51317 | BHLR transcritpion |
| 33 | 10 | 22.7 | 6 2 I49421 | lamnin B1 - waste |
| 34 | 10 | 22.7 | 7 2 S33244 | neuromodulatory pe |
| 35 | 10 | 22.7 | 7 2 S33245 | neuromodulatory pe |
| 36 | 10 | 22.7 | 7 2 A58718 | carnocin U149 - Ca |
| 37 | 10 | 22.7 | 8 2 B24749 | neuropeptide B - b |
| 38 | 10 | 22.7 | 8 2 S22428 | chitin-binding pro |
| 39 | 10 | 22.7 | 8 2 A46306 | spasmodic toxin |
| 40 | 10 | 22.7 | 8 2 A28719 | thymic humoral fac |
| 41 | 10 | 22.7 | 8 2 A42057 | fibroblast growth |
| 42 | 10 | 22.7 | 8 2 A25836 | L-serine amonia-1 |
| 43 | 10 | 22.7 | 9 2 B45796 | dihydrolipeamide S |
| 44 | 10 | 22.7 | 9 2 A91466 | oxytocin - hippo |
| 45 | 10 | 22.7 | 9 2 A92774 | oxytocin - spotted |

ALIGNMENTS

RESULT 1
PT0315
Ig heavy chain CRD3 region (clone 6-109) - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C/Accession: PT0315
R/Yamada, M.; Masserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
U. Exp. Med. 173, 395-407, 1991
A/Title: Preferential utilization of specific immunoglobulin heavy chain diversity and
A/Reference number: PT0222; MUID:91108337; PMID:1899102
A/Accession: PT0315
A/Molecule type: DNA
A/Residues: 1-9 <YAM>
A/Keywords: heterotetramer; immunoglobulin

Query Match 38.6%; Score 17; DB 2; Length 9;
Best Local Similarity 42.9%; Pred. No. 2.8e+05;
Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 EMEFLQIY 9
DB 2 ERFLESP 8

RESULT 2
A27696
contraction-inhibiting peptide I - blue mussel
C/Species: Mytilus edulis (blue mussel)
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C/Accession: A27696
R/Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A/Title: Structures and actions of Mytilus inhibitory peptides.
A/Reference number: A90142; MUID:86240357; PMID:3377776
A/Accession: A27696
A/Molecule type: protein
A/Residues: 1-6 <HIR>
C/Keywords: amidated carboxyl end
F/6/Modified site: amidated carboxyl end (Val) #status experimental

Query Match 34.1%; Score 15; DB 2; Length 6;
Best Local Similarity 60.0%; Pred. No. 2.8e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMFL 6
DB 2 SPMFLV 6

RESULT 3
S08606

hypothetical protein 2 estrogen receptor 5'-region - chicken

C:Species: Gallus gallus (chicken)
C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 18-Aug-2000
C/Accession: S08606
R/Krust, A.; Green, S.; Argos, P.; Kumar, V.; Walter, P.; Bornert, J.M.; Chambon, P.
EMBO J. 5, 891-897, 1986
A>Title: The chicken oestrogen receptor sequence: homology with v-erbA and the human oes
A/Reference number: S07192; MUID:86247578; PMID:3755102
A/Accession: S08606
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-7 <R>
A/Cross-references: EMBL:X03805; NID:663378; PIDN:CAA27432.1; PID:9584450
C/Superfamily: unassigned leader peptides

Query Match 34.1%; Score 15; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 MFL 6
: : :
DB 1 MFL 3

RESULT 4
152974
seminal vesicle protein IV - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C/Accession: I52974
R/Teng, C.T.; Harris, S.E.
DNA 2, 105-111, 1983
A>Title: The seminal vesicle secretion IV gene: detection of SI nuclease-sensitive sites
A/Reference Number: I52974; MUID:83261204; PMID:6307619
A/Accession: I52974
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-9 <R>
A/Cross-references: GB:M27324; NID:9207124; PIDN:AAA63501.1; PID:9207125
C/Genetics:
A/Gene: SVSIV

Query Match 34.1%; Score 15; DB 2; Length 9;
Best Local Similarity 33.3%; Pred. No. 2.8e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEMFL 6
: : :
DB 3 LTRFL 8

RESULT 5
PT0225
Ig heavy chain CDR3 region (clone 1-103) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C/Accession: PT0225
R/Yamada, M.; Masegama, R.; Reichard, B.A.; Shane, S.; Caton, A.D.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A>Title: Preferential utilization of specific immunoglobulin heavy chain diversity and J
A/Reference number: PT0222; MUID:91108337; PMID:1899102
A/Accession: PT0225
A/Molecule type: DNA
A/Residues: 1-9 <YAM>
A/Experimental source: B lymphocyte
C/Keywords: heterotrimer; immunoglobulin

Query Match 29.5%; Score 13; DB 2; Length 9;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5
: : :
DB 1

DB 1 DMF 3

RESULT 6
149406
bone gla protein - western wild mouse (fragment)
C:Species: Mus spreus (western wild mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C/Accession: I49406
R/Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J
Mamm. Genome 5, 349-355, 1994
A>Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
A/Reference number: I48934; MUID:94319082; PMID:8043949
A/Accession: I49406
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-9 <RES>
A/Cross-references: EMBL:U05695; NID:9497015; PIDN:AAB60465.1; PID:9642828

Query Match 29.5%; Score 13; DB 2; Length 9;
Best Local Similarity 40.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 FLOIY 9
: : :
DB 1 YKRIY 5

RESULT 7
154357
schwannomin - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C/Accession: I54357
R/Huynh, D.P.; Nechiporuk, T.; Pulst, S.
Hum. Mol. Genet. 3, 1075-1079, 1994
A>Title: Alternative transcripts in the mouse neurofibromatosis type 2 (NF2) gene are c
A/Reference number: I54357; MUID:95072570; PMID:7981675
A/Accession: I54357
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-4 <RES>
A/Cross-references: GB:I28838; NID:9454836; PIDN:AAA57150.1; PID:9601923
C/Genetics:
A/Gene: NF2

Query Match 27.3%; Score 12; DB 2; Length 4;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISEM 4
: : :
DB 1 VPEM 4

RESULT 8
B27696
contraction-inhibiting peptide II - blue mussel)
C:Species: Mytilus edulis (blue mussel)
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C/Accession: B27696
R/Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Munekata, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A>Title: Structures and actions of Mytilus inhibitory peptides.
A/Reference number: A90142; MUID:88240357; PMID:3377776
A/Accession: B27696
A/Molecule type: Protein
A/Residues: 1-6 <HR>
C/Keywords: amidated carboxyl end
F/6/Modified site: amidated carboxyl end (Val) #status experimental

Query Match 27.3%; Score 12; DB 2; Length 6;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;

Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFL 6
|:
Db 4 MFV 6

RESULT 9

B39127
Phosphotransferase system enzyme II (EC 2.7.1.69) - Escherichia coli (fragment)
C/Species: Escherichia coli
C/Date: 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 08-Oct-1999
C/Accession: B39127
R/Hardisty, C.; Ferran, C.; DiRienzo, J.M.
J. Bacteriol. 173, 449-456, 1991
A/Title: Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of sc rin.
A/Reference number: A39127; MUID:91100329; PMID:1846143
A/Accession: B39127
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-7 <HMR>
A/Cross-references: GB:M38416; NID:g155142; PIDN:AAA98418.1; PID:g155144
C/Keywords: phosphotransferase

Query Match 27.3%; Score 12; DB 2; Length 7;
Best Local Similarity 75.0%; Pred. No. 2.8e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 FLOI 8
|:
Db 3 FEOI 6

RESULT 10
A41117
acetylcholinesterase (EC 3.1.1.7), venom - Asian cobra (fragment)
C/Species: Naja naja oxiata (Asian cobra, Oxiu cobra)
C/Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 23-Jun-1993
C/Accession: A41117
R/Kreienkamp, H.J.; Weise, C.; Raba, R.; Aavikaar, A.; Hucho, F.
Proc. Natl. Acad. Sci. U.S.A. 88, 6117-6121, 1991
A/Title: Anionic subsites of the catalytic center of acetylcholinesterase from Torpedo a
A/Reference number: A41117; MUID:91296772; PMID:2068091
A/Accession: A41117
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-8 <KRE>
C/Keywords: carboxylic ester hydrolase

Query Match 27.3%; Score 12; DB 2; Length 8;
Best Local Similarity 50.0%; Pred. No. 2.8e+05;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SEMP 5
|:
Db 2 AEMW 5

RESULT 11
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C/Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C/Accession: T13818
R/Delapierre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A/Title: The main features of the craniate mitochondrial DNA between the NDI and the COI
A/Reference number: Z17775; MUID:97398704; PMID:9254918
A/Accession: T13818
A/Status: preliminary; translated from GB/EMBL/DBD
A/Molecule type: DNA
A/Residues: 1-8

A/Cross-references: EMBL:Y09527; NID:92340019; PIDN:CAA70718.1; PID:92340022
C/Genetics:
A/Genome: mitochondrion
A/Note: COI
C/Keywords: mitochondrion

Query Match 27.3%; Score 12; DB 2; Length 8;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 MFL 6
|:
Db 1 MYL 3

RESULT 12
D58503
translation elongation factor EF-Tu - unidentified bacterium (fragment)
C/Species: unidentified bacterium
C/Date: 07-Feb-1997 #sequence_revision 07-Feb-1997 #text_change 28-May-1999
C/Accession: D58503
R/Binette, J.P.; Binette, M.B.
submitted to the Protein Sequence Database, October 1996
A/Description: The proteins of kidney and gallbladder stones.
A/Reference number: A58501
A/Accession: D58503
A/Status: preliminary
A/Molecule type: Protein
A/Residues: 1-9 <BIN>
A/Experimental source: human bile and stones
C/Superfamily: translation elongation factor Tu; translation elongation factor Tu homol.
C/Keywords: GTP binding

Query Match 27.3%; Score 12; DB 2; Length 9;
Best Local Similarity 66.7%; Pred. No. 2.8e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 QTY 9
|:
Db 5 QFY 7

RESULT 13
B41978
CalliPMPamide 2 - bluebottle fly (Calliphora vomitoria)
C/Species: Calliphora vomitoria
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C/Accession: B41978
R/Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorp
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A/Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (des)
A/Reference number: A41978; MUID:92196111; PMID:1549595
A/Accession: B41978
A/Status: preliminary
A/Molecule type: Protein
A/Residues: 1-9 <DUV>
C/Keywords: amidated carboxyl end; neuropeptide
F/9/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.3%; Score 12; DB 2; Length 9;
Best Local Similarity 33.3%; Pred. No. 2.8e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMPLO 7
|:
Db 3 SODPWR 8

RESULT 14
C41978
CalliPMPamide 3 - bluebottle fly (Calliphora vomitoria)
C/Species: Calliphora vomitoria
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C:/Accession: C41978
 R:/Dive, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorpe
 Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
 A:/Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
 A:/Reference number: A41978; MUID:92196111; PMID:1549595
 A:/Accession: C41978
 A:/Status: preliminary
 A:/Molecule type: protein
 A:/Residues: 1-9 <DUV>
 C:/Keywords: amidated carboxyl end; neuropeptide
 F:/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.3%; Score 12; DB 2; Length 9;
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;
 Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SEMPLQ 7
 | : | : | :
 Db 3 SODPMR 8

RESULT 15

A35779
 neuropeptide Antho-RNamide - sea anemone (Anthopleura elegantissima)
 C:/Species: Anthopleura elegantissima
 C:/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 08-Dec-1995
 C:/Accession: A35779
 R:/Grimmelikhuizen, C.J.P.; Rinehart, K.L.; Jacob, E.; Graff, D.; Reinscheid, R.K.; Noth
 Proc. Natl. Acad. Sci. U.S.A. 87, 5410-5414, 1990
 A:/Title: Isolation of L-3-phenyllactyl-Leu-Arg-Asn-NH2 (Antho-RNamide), a sea anemone ne
 A:/Reference number: A35779; MUID:90319122; PMID:1973541
 A:/Accession: A35779
 A:/Molecule type: protein
 A:/Residues: 1-4 <GR1>
 C:/Comment: The L-3-phenyllactyl blocking group probably arises from an amino-terminal ph
 C:/Keywords: amidated carboxyl end; neuropeptide; phenyllactylation
 F:/Modified site: L-3-phenyllactic acid (Phe) #status experimental
 F:/Modified site: amidated carboxyl end (Asn) #status experimental

Query Match 25.0%; Score 11; DB 2; Length 4;
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 FLQ 7
 | : | : | :
 Db 1 FLR 3

Search completed: May 1, 2003, 23:24:24
 Job time : 16 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 13:22:52 ; Search time 396 Seconds
(without alignments)
8939,751 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gaattccctgcgcttgtaa.....tcgcctctgcgctgagctgg 1572

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq, 101002.*
1: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-----------------------------|
| 1 | 1572 | 100.0 | 1572 | 22 | AA500585 Human MUC1 DNA seq |
| 2 | 1481.2 | 94.2 | 1721 | 24 | ABL68032 Ovary cancer relat |
| 3 | 1481.2 | 94.2 | 1721 | 24 | ABL68032 Human ovarian rela |
| 4 | 1412.2 | 89.8 | 1804 | 24 | ABL67539 Thyroid cancer rel |
| 5 | 1342 | 85.4 | 1428 | 24 | ABL60159 Human MUC1 encodin |
| 6 | 1313 | 83.5 | 1452 | 13 | AAQ29276 Encodes transmembr |
| 7 | 1257 | 80.0 | 1527 | 19 | AAV48329 MIM10C1 gene. Ho |
| 8 | 1213.4 | 77.2 | 1737 | 21 | UBIQUITIN-E. coli |
| 9 | 1210.2 | 77.0 | 1371 | 21 | AA000388 Human Mucin 1 (MUC |

| | | | | | |
|----|--------|------|------|----|----------|
| 10 | 1190.2 | 75.7 | 4139 | 24 | ABN55623 |
| 11 | 1190.2 | 75.7 | 4139 | 24 | ABE67071 |
| 12 | 1190.2 | 75.7 | 4139 | 24 | ABE67544 |
| 13 | 1190.2 | 75.7 | 4139 | 24 | ABE68884 |
| 14 | 1190.2 | 75.7 | 4139 | 24 | ABK09797 |
| 15 | 1056 | 67.2 | 1074 | 13 | AAQ24679 |
| 16 | 993.4 | 63.2 | 1320 | 16 | AAI29190 |
| 17 | 978 | 62.2 | 1355 | 21 | AACT7625 |
| 18 | 750 | 47.7 | 6286 | 24 | ABL60158 |
| 19 | 750 | 47.7 | 8181 | 24 | ABL41001 |
| 20 | 738.4 | 47.0 | 1709 | 13 | AAQ29277 |
| 21 | 661.4 | 42.1 | 822 | 17 | AAI10677 |
| 22 | 661.4 | 42.1 | 849 | 17 | AAI10678 |
| 23 | 638.2 | 40.6 | 1820 | 21 | AAQ00392 |
| 24 | 635.4 | 40.6 | 1820 | 21 | AAQ00392 |
| 25 | 607.4 | 38.6 | 768 | 17 | AAI10679 |
| 26 | 607.4 | 38.6 | 795 | 17 | AAI10680 |
| 27 | 587.2 | 37.4 | 723 | 17 | AAI10681 |
| 28 | 587.2 | 37.4 | 750 | 17 | AAI10682 |
| 29 | 540 | 34.4 | 558 | 13 | AAQ24678 |
| 30 | 540 | 34.4 | 558 | 13 | AAQ24680 |
| 31 | 494.4 | 31.5 | 497 | 24 | ABK09657 |
| 32 | 481.4 | 30.6 | 1331 | 13 | AAQ24681 |
| 33 | 459 | 29.2 | 5660 | 24 | ABL34357 |
| 34 | 429.2 | 27.3 | 459 | 12 | AAQ12772 |
| 35 | 366 | 23.3 | 891 | 21 | AAQ00391 |
| 36 | 362.8 | 23.1 | 525 | 21 | AAQ00385 |
| 37 | 328.8 | 20.9 | 1763 | 12 | AAQ12771 |
| 38 | 328.4 | 20.9 | 330 | 22 | AAE68091 |
| 39 | 328.4 | 20.9 | 330 | 22 | ABK88002 |
| 40 | 285.8 | 18.2 | 1567 | 12 | AAQ12773 |
| 41 | 279 | 17.7 | 285 | 21 | AAQ00387 |
| 42 | 277 | 17.6 | 651 | 21 | AAQ00393 |
| 43 | 253.6 | 16.1 | 309 | 10 | AAAN0579 |
| 44 | 237 | 15.1 | 237 | 24 | ABK09556 |
| 45 | 235.4 | 15.0 | 5660 | 24 | ABL34356 |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AA500585 | AA500585 standard; DNA; 1572 BP. |
| ID | |
| XX | |
| AC | AA500585; |
| XX | |
| DT | 12-SEP-2001 (first entry) |
| XX | |
| DE | Human MUC1 DNA sequence. |
| XX | |
| KW | Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I; |
| KW | glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine; |
| KW | cancer gene therapy; diagnosis; treatment; inflammatory disorder; de; |
| KW | organ transplant rejection; graft versus host disease. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PH | Key |
| FT | CDS |
| FT | 58..1545 |
| FT | Location/Qualifiers |
| FT | /*tag= a |
| FT | /product= "Human MUC1" |
| XX | |
| XX | MO200118035-A2. |
| XX | |
| PD | 15-MAR-2001. |
| XX | |
| PF | 07-SEP-2000; 2000MO-EP08761. |
| XX | |
| PR | 08-SEP-1999; 99GP-0021242. |
| PR | 10-SEP-1999; 99EP-0402237. |
| PR | 03-MAR-2000; 2000US-0187215. |
| XX | |

Gene #2121 used to
Thyroid cancer rel
Thyroid cancer rel
Kidney cancer rela
Human ovarian tumo
H23-ETA-T antigen
Glycoprotein 39 3'
Human cancer assoc
Human MUC1 encodin
Human neuquelin 2
Encodes secreted f
Mucin-derived prot
Mucin-derived prot
Ubiquitin-E. coli
Human Mucin 1 (MUC
Mucin-derived prot
Mucin-derived prot
Mucin-derived prot
Mucin-derived prot
H23-ETA-T antigen
H23-ETA-S antigen
Human ovarian tumo
H23-ETA-S antigen
Human immune syste
Human polymorphic
Ubiquitin-E. coli
Human Mucin 1 (MUC
Human polymorphic
Human lung tumour
cDNA encoding clon
Human polymorphic
Human Mucin 1 (MUC
Ubiquitin-E. coli
pDPF.3 cDNA insert
Human ovarian tumo
Human immune syste

PA (TRGE) TRANSGENE SA.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PI Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;
PI Acres B, Thomas M;
XX
XX MPI: 2001-235187/24.
DR P-PSDB; AAU00539.
XX
XX
PT New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
XX Claim 1; Page 66-68; 81pp; English.
XX
XX The sequence represents a polynucleotide which encodes MUC1 polypeptide.
CC derivative antigenic peptides of this protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumours.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of a MHC class I restricted T cell response to a MUC1
CC derivative, where the presence of the MHC class I restricted T cell
CC response indicates that the host has cancer. Measurement of the level of
CC MHC class I restricted T cell response is also useful to monitor the
CC severity of cancer, a larger response indicating a more severe cancer.
CC MUC1 derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer, gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
SQ Sequence 1572 BP; 312 A; 567 C; 364 G; 329 T; 0 other;

Query Match 100.0%; Score 1572; DB 22; Length 1572;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCAACCCATTTCACCAACCACTG 60
Db 1 GAATTCCTGCTGCTTGAATCTGTTCTGCCCCCTCCCAACCCATTTCACCAACCACTG 60
QY 61 AACCGGGGACACCGATCTCTTCTTCTGCTGCTGCTCTCTCAAGTCTTACAGTTT 120
Db 61 AACCGGGGACACCGATCTCTTCTTCTGCTGCTGCTCTCTCAAGTCTTACAGTTT 120
QY 121 AAGGTTTGTGTATGCAAGCTTACCCAGGTGAGAAAAGAGATTTCGGTACCCAG 180
Db 121 AAGGTTTGTGTATGCAAGCTTACCCAGGTGAGAAAAGAGATTTCGGTACCCAG 180
QY 181 AAGAGTTCAAGTCCCAAGCTCTAAGTGAAGAAATCTGTGATATGACAGAGGATCT 240
Db 181 AAGAGTTCAAGTCCCAAGCTCTAAGTGAAGAAATCTGTGATATGACAGAGGATCT 240
QY 241 TCCAGCCACAGACCCCGGTTTCAGGCTCTCTCAACATGAGGAGACAGATGTCAGCC 300
Db 241 TCCAGCCACAGACCCCGGTTTCAGGCTCTCTCAACATGAGGAGACAGATGTCAGCC 300
QY 301 CGGGGCAAGGAAACAGTTTCAAGTTCAAGTCCCACTTGGGAGACAGATGTCAGCTG 360
Db 301 CGGGGCAAGGAAACAGTTTCAAGTTTCAAGTCCCACTTGGGAGACAGATGTCAGCTG 360
QY 361 CGAGTCAACAGGACCGCTTGAGCTTCAACACCCCGCAGCCCAAGATGTCAGCTGAGCC 420
Db 361 CGAGTCAACAGGACCGCTTGAGCTTCAACACCCCGCAGCCCAAGATGTCAGCTGAGCC 420
QY 421 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 480
Db 421 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 480
QY 481 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 540

Db 481 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 540
QY 541 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 600
Db 541 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 600
QY 601 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 660
Db 601 CCGGACAAAGAGCCCGGGCTTCAACCGCCCGCGGCCCAAGTGTACCTCGGCC 660
QY 661 TCAAGCTTGAATGAGCTCAAGTTCTTACTGTGTGCAAAAGGACCTTTCGACAGGCT 720
Db 661 TCAAGCTTGAATGAGCTCAAGTTCTTACTGTGTGCAAAAGGACCTTTCGACAGGCT 720
QY 721 ACCACAAACCCAGCCAGACAGACATCCACCCAGATTCACGACCACTGTGATCT 780
Db 721 ACCACAAACCCAGCCAGACAGACATCCACCCAGATTCACGACCACTGTGATCT 780
QY 781 CTTACCAACCTTGGCAGACGATACAGACAAAGATGATGATGATGATGATGATGATGATG 840
Db 781 CTTACCAACCTTGGCAGACGATACAGACAAAGATGATGATGATGATGATGATGATGATG 840
QY 841 GTACTCTCTTCACTCTCTTCAATCAAGACATTTCTCCAGTTGTCTACTGGGATCTCT 900
Db 841 GTACTCTCTTCACTCTCTTCAATCAAGACATTTCTCCAGTTGTCTACTGGGATCTCT 900
QY 901 TTTCTTTTCTGCTTTTTCATTTCAATTCAAACCTTCAGTTTATTCCTCTCTGAGATCCC 960
Db 901 TTTCTTTTCTGCTTTTTCATTTCAATTCAAACCTTCAGTTTATTCCTCTCTGAGATCCC 960
QY 961 AGACACGACTCTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTCAGATTTAT 1020
Db 961 AGACACGACTCTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTCAGATTTAT 1020
QY 1021 AAACAAAGGGGTTTTTCTGGGCTCTTCAATATTAAGTTCAAGCCAGATCTGTGTGTA 1080
Db 1021 AAACAAAGGGGTTTTTCTGGGCTCTTCAATATTAAGTTCAAGCCAGATCTGTGTGTA 1080
QY 1081 CAATTGACTCTGGGCTTCCAGAAAGGATCAATGATGATGATGATGATGATGATGATGATG 1140
Db 1081 CAATTGACTCTGGGCTTCCAGAAAGGATCAATGATGATGATGATGATGATGATGATGATG 1140
QY 1141 AATCAGTATTAATAAGGAAGAGCTCTGATATTAACCTGACATCTCAGAGCTGAGG 1200
Db 1141 AATCAGTATTAATAAGGAAGAGCTCTGATATTAACCTGACATCTCAGAGCTGAGG 1200
QY 1201 AGTCATGTGCAATTTCTTTCTGCTGCAAGTCTGGGCTGGGGTGCAGAGCTGGGCAATC 1260
Db 1201 AGTCATGTGCAATTTCTTTCTGCTGCAAGTCTGGGCTGGGGTGCAGAGCTGGGCAATC 1260
QY 1261 GCGCTGCTGGT 1320
Db 1261 GCGCTGCTGGT 1320
QY 1321 GCTGTCTGAGTGCCTCCGAAAGAACTACGAGGAGCTGAGACATCTTTCAGGCCGAGAT 1380
Db 1321 GCTGTCTGAGTGCCTCCGAAAGAACTACGAGGAGCTGAGACATCTTTCAGGCCGAGAT 1380
QY 1381 ACCTACATCTTAAGAGAGATCCCAACCTACACACCAATGAGGATGATGATGATGATGATG 1440
Db 1381 ACCTACATCTTAAGAGAGATCCCAACCTACACACCAATGAGGATGATGATGATGATGATG 1440
QY 1441 AGCAGTACCGATCGTACGCTCTTATGAGAAAGTTTCTGACAGTATGATGATGATGATGATG 1500
Db 1441 AGCAGTACCGATCGTACGCTCTTATGAGAAAGTTTCTGACAGTATGATGATGATGATGATG 1500
QY 1501 TCTTACAAACCAAGAGAGTGCAGGCACTTCTGCAACTTGTAGAGGAGACGTGCGCTTC 1560
Db 1501 TCTTACAAACCAAGAGAGTGCAGGCACTTCTGCAACTTGTAGAGGAGACGTGCGCTTC 1560
QY 1561 TTAGCTGATGG 1572

| | | | |
|----|------|---|------|
| Qy | 541 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 600 |
| Db | 601 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 660 |
| Qy | 601 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 660 |
| Db | 661 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 720 |
| Qy | 661 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 720 |
| Db | 721 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 780 |
| Qy | 721 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 780 |
| Db | 781 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 840 |
| Qy | 781 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 840 |
| Db | 841 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 900 |
| Qy | 841 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 900 |
| Db | 901 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 960 |
| Qy | 901 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 960 |
| Db | 961 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1020 |
| Qy | 961 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1020 |
| Db | 1021 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1080 |
| Qy | 1021 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1080 |
| Db | 1081 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1140 |
| Qy | 1081 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1140 |
| Db | 1141 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1200 |
| Qy | 1141 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1200 |
| Db | 1201 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1260 |
| Qy | 1201 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1260 |
| Db | 1261 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1320 |
| Qy | 1261 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1320 |
| Db | 1321 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1380 |
| Qy | 1321 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1380 |
| Db | 1381 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1440 |
| Qy | 1381 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1440 |
| Db | 1441 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1500 |
| Qy | 1441 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1500 |
| Db | 1501 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1560 |
| Qy | 1501 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1560 |
| Db | 1561 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1620 |
| Qy | 1561 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1620 |
| Db | 1621 | CCGACACACAGCGCGGCGCTTGGGCGTCCAGCGGCGCTTCAGACCAAGGTACATCTCGGCG | 1680 |

| ID | ABK09743 | standard; cDNA; 1721 BP. |
|----|--|--------------------------|
| XX | ABK09743 | |
| AC | ABK09743; | |
| DT | 14-MAR-2002 | (first entry) |
| DE | Human ovarian tumour protein encoding cDNA #276. | |
| XX | | |
| XX | Human; ovarian tumour protein; cancer; cytostatic; immunostimulant; ss; | |
| KW | gene therapy; CD4+ T cell; CD8+ T cell; PCR primer. | |
| OS | Human sapiens. | |
| XX | | |
| XX | MO200190154-A2. | |
| FN | | |
| XX | | |
| FD | 29-NOV-2001. | |
| PD | | |
| XX | | |
| PF | 23-MAY-2001; 2001WO-US16895. | |
| XX | | |
| PP | 24-MAY-2000; 2000US-207107P. | |
| PR | 13-JUN-2000; 2000US-211457P. | |
| PR | 21-JUN-2000; 2000US-213673P. | |
| PR | 03-AUG-2000; 2000US-223288P. | |
| PR | 01-MAR-2001; 2001US-272790P. | |
| XX | | |
| XX | (COR1-) CORIXA CORP. | |
| PA | | |
| XX | | |
| XX | Xu J, Mitcham JI, Harlocker SL, Dillon DC, Secrist H, Lodes MJ; | |
| P1 | Algate PA, Fling SP, Mannion J, Benson DR, Carter D; | |
| P1 | | |
| XX | WPI: 2002-097641/73. | |
| DR | | |
| XX | | |
| PT | New isolated polynucleotide encoding polypeptide comprising portion of | |
| PT | ovarian tumour protein, useful for detection, diagnosis and therapy of | |
| PT | human ovarian cancer - | |
| PS | | |
| XX | Claim 1; Page 222; 285pp; English. | |
| XX | | |
| XX | The invention relates to an isolated polynucleotide encoding a | |
| CC | polypeptide comprising a portion of an ovarian tumour protein. The | |
| CC | sequences of the invention are useful for stimulating an immune response | |
| CC | and for treating ovarian cancer in a patient. An antigen presenting cell | |
| CC | that expresses the sequences is useful for treating ovarian cancer by | |
| CC | incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells | |
| CC | can then be proliferated and administered to the patient to inhibit the | |
| CC | development of cancer. The DNA sequences are useful as probes or primers | |
| CC | for nucleic acid hybridization, to direct expression of a polypeptide in | |
| CC | appropriate host cells. Detecting the presence of a cancer in a patient | |
| CC | involves obtaining a biological sample from the patient, detecting the | |
| CC | amount of protein that binds to the agent, comparing the amount of | |
| CC | protein to a predetermined cut-off value and determining the presence of | |
| CC | cancer. Sequences ABK09464-ABK09802 represent PCR primers and cDNA | |
| CC | molecules encoding ovarian tumour proteins of the invention. | |
| XX | | |
| XX | Sequence 1721 BP; 339 A; 634 C; 395 G; 353 T; 0 other; | |
| XX | | |
| XX | Query Match 94.2%; Score 1481.2; DB 24; Length 1721; | |
| XX | Best Local Similarity 95.5%; Pred. No. 1.6e-311; | |
| XX | Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1 | |
| OY | 1 GAATTCCTGGCTGCTGAATCTGTTCTGCCCCCTCCCAACCATTTACACACACATG 60 | |
| DB | 1 GAATTCCTGGCTGCTGAATCTGTTCTGCCCCCTCCCAACCATTTACACACACATG 60 | |
| OY | 61 ACAAAGGACACCAAGTCTCCCTTCTTCTGCTGCTGCTCTCAACAGTGTACAGTTGTT 120 | |
| DB | 61 ACAAAGGACACCAAGTCTCCCTTCTTCTGCTGCTGCTCTCAACAGTGTACAGTTGTT 120 | |
| OY | 121 ACAAAGTCTGGTATGATGCAAGCTTCAACCAAGTGTGAGAAAGAGACTTGGCTACAG 180 | |

QY 838 AGGTAACCTCTCACTCTCAATCAAGACACTTCTCCCAAGTGTCTACTGGGATC 897
 DB 721 AGGTAACCTCTCACTCTCAATCAAGACACTTCTCCCAAGTGTCTACTGGGATC 780
 QY 898 TCTTCTTTTCTCTGCTCTTCAATTTGAAAGCTCGAGTTATTTCTCTGGAAGAT 957
 DB 781 TCTTCTTTTCTCTGCTCTTCAATTTGAAAGCTCGAGTTATTTCTCTGGAAGAT 840
 QY 958 CCGAGCACCGACTATCAACAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 1017
 DB 841 CCGAGCACCGACTATCAACAGAGCTGACAGACATTTTGAAATGTTTTCAGATT 900
 QY 1018 TATTAACAGAGGAGGTTTCTGAGGCTCTCAATATTAAGTTAGGCGCAGATCTGTG 1077
 DB 901 TATTAACAGAGGAGGTTTCTGAGGCTCTCAATATTAAGTTAGGCGCAGATCTGTG 960
 QY 1078 GPACAAATTGACTGCTGGCTTCGAGAAAGTACCAATCAATGTCCAGACTGGAAGACAG 1137
 DB 961 GPACAAATTGACTGCTGGCTTCGAGAAAGTACCAATCAATGTCCAGACTGGAAGACAG 1020
 QY 1138 TTCAATCAGTATTAACGGAAGACAGCTCTGATATPACCTGACATCTCGACGTGAC 1197
 DB 1021 TTCAATCAGTATTAACGGAAGACAGCTCTGATATPACCTGACATCTCGACGTGAC 1080
 QY 1198 GTGAGTCATGTCATTTCTTTCTCTGCGCCAGTCTGAGGCTGAGGCTGAGGCTG 1257
 DB 1081 GTGAGTCATGTCATTTCTTTCTCTGCGCCAGTCTGAGGCTGAGGCTGAGGCTG 1140
 QY 1258 ATCGGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
 DB 1141 ATCGGCTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 QY 1318 TTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
 DB 1201 TTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 QY 1378 GATACCTACCATCTTATGAGCGAGTACCCCACTACCAACCATAGGCTGATGTGCC 1437
 DB 1261 GATACCTACCATCTTATGAGCGAGTACCCCACTACCAACCATAGGCTGATGTGCC 1320
 QY 1438 CCGAGCAGTACCGACTGATGAGGCTTCTGAGAGTATGAGTGTGAGCAGC 1497
 DB 1321 CCGAGCAGTACCGACTGATGAGGCTTCTGAGAGTATGAGTGTGAGCAGC 1380
 QY 1498 CTCTCTTACCAACCCAGACAGTGGACCACTTCTGCACTGTAG 1545
 DB 1381 CTCTCTTACCAACCCAGACAGTGGACCACTTCTGCACTGTAG 1428
 RESULT 6
 AAQ29276 ID AAQ29276 standard; cDNA; 1452 BP.
 AC AAQ29276:
 DT 06-NOV-1992 (first entry)
 XX
 DE Encodes transmembrane form of H23-ETA antigen.
 XX ETA-T; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KM malignant tumour; breast cancer; ss.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH sig_peptide 58..120
 FT /tag= a
 FT mat_peptide 121..1422
 FT /tag= b
 FT misc_feature 439..498
 FT /tag= c
 FT /note= "epitope recognised by Mab H23, can be
 present in 1-80 tandem repeats"

FT variation 457..459
 FT /tag= d
 FT /note= "either an Ala or a Pro codon"
 FT variation 487..489
 FT /tag= e
 FT /note= "either a Thr or an Asn codon"
 FT variation 496..498
 FT /tag= f
 FT /note= "either a Ala or an Pro codon"
 PN MO9207000-A.
 PD 30-APR-1992.
 PF 23-OCT-1991; 91WO-FR00835.
 PR 23-OCT-1990; 90FR-0013101.
 XX (TRGE) TRANSGENE SA.
 DN Chamhon P, Hareuveni M, Kieny MP, Lathe R,
 PL WPI; 1992-167097/20.
 DR P-PSDB; AAR23973.
 XX Compens. conty. polypeptide antigen recognised by antibody H23 -
 PT for treatment of mammary tumours, also for pox virus compens. for
 PT use in vaccines
 PS Claim 3; Page 15-18; 29pp; French.
 XX
 CC The tumour antigen recognised by antibody H23 is aberrantly
 CC expressed in epithelial cells from cancerous mammary tissue in about
 CC 90 per cent of breast cancer cases; in a normal individual
 CC expression is negligible. The antigen exists in two forms:
 CC transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high
 CC degree of polymorphism. A 20 amino acid subunit in ETA can be
 CC tandemly repeated up to 80 times. From one subunit to the next, 1 to
 CC 3 amino acids can differ (= "variation" in the features table). DNA
 CC coding for immunogenic fragments of ETA can be inserted into
 CC e.g. vaccinia viruses for treatment of mammary tumours.
 CC See also AAQ24678-Q24681, AAQ29277 and AAR23974-R23981.
 XX
 SO Sequence 1452 BP; 298 A; 495 C; 329 G; 321 T; 9 other:
 Query Match 83.5%; Score 1313; DB 13; Length 1452;
 Best Local Similarity 91.8%; Pred. No. 4.4e-275;
 Matches 1443; Conservative 0; Mismatches 9; Indels 120; Gaps 1;
 QY 1 GAATTCCTGGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
 DB 1 GAATTCCTGGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
 QY 61 ACACCGGGACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 DB 61 ACACCGGGACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 QY 121 ACAGTTCGTCTATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTACCCAG 180
 DB 121 ACAGTTCGTCTATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGGCTACCCAG 180
 QY 181 AGAAGTTAGTCCCACTCTACTGAGAAAGATCTGTGATATGACAGCAGAGGTAATC 240
 DB 181 AGAAGTTAGTCCCACTCTACTGAGAAAGATCTGTGATATGACAGCAGAGGTAATC 240
 QY 241 TCCAGCAGACAGCCCGGTTCAGGCTCTCCACACTGAGGAGACAGATGTCACTTGAGCC 300
 DB 241 TCCAGCAGACAGCCCGGTTCAGGCTCTCCACACTGAGGAGACAGATGTCACTTGAGCC 300
 QY 301 CCGGCAACGAAACAGCTTACAGTTCAGCTGCACTGAGGAGACAGATGTCACTTGAGCTC 360
 DB 301 CCGGCAACGAAACAGCTTACAGTTCAGCTGCACTGAGGAGACAGATGTCACTTGAGCTC 360

[illegible]

| Db | Accession | Score | Length | Mismatches | Indels | Gaps |
|------|-----------|---------|--------|------------|--------|------|
| Db1 | 1321 | 100.0% | 1527 | 0 | 0 | 0 |
| Db2 | 1501 | 99.9% | 1527 | 1 | 0 | 0 |
| Db3 | 1381 | 99.8% | 1527 | 2 | 0 | 0 |
| Db4 | 1561 | 99.7% | 1527 | 3 | 0 | 0 |
| Db5 | 1441 | 99.6% | 1527 | 4 | 0 | 0 |
| Db6 | 1501 | 99.5% | 1527 | 5 | 0 | 0 |
| Db7 | 1381 | 99.4% | 1527 | 6 | 0 | 0 |
| Db8 | 1561 | 99.3% | 1527 | 7 | 0 | 0 |
| Db9 | 1441 | 99.2% | 1527 | 8 | 0 | 0 |
| Db10 | 1501 | 99.1% | 1527 | 9 | 0 | 0 |
| Db11 | 1381 | 99.0% | 1527 | 10 | 0 | 0 |
| Db12 | 1561 | 98.9% | 1527 | 11 | 0 | 0 |
| Db13 | 1441 | 98.8% | 1527 | 12 | 0 | 0 |
| Db14 | 1501 | 98.7% | 1527 | 13 | 0 | 0 |
| Db15 | 1381 | 98.6% | 1527 | 14 | 0 | 0 |
| Db16 | 1561 | 98.5% | 1527 | 15 | 0 | 0 |
| Db17 | 1441 | 98.4% | 1527 | 16 | 0 | 0 |
| Db18 | 1501 | 98.3% | 1527 | 17 | 0 | 0 |
| Db19 | 1381 | 98.2% | 1527 | 18 | 0 | 0 |
| Db20 | 1561 | 98.1% | 1527 | 19 | 0 | 0 |
| Db21 | 1441 | 98.0% | 1527 | 20 | 0 | 0 |
| Db22 | 1501 | 97.9% | 1527 | 21 | 0 | 0 |
| Db23 | 1381 | 97.8% | 1527 | 22 | 0 | 0 |
| Db24 | 1561 | 97.7% | 1527 | 23 | 0 | 0 |
| Db25 | 1441 | 97.6% | 1527 | 24 | 0 | 0 |
| Db26 | 1501 | 97.5% | 1527 | 25 | 0 | 0 |
| Db27 | 1381 | 97.4% | 1527 | 26 | 0 | 0 |
| Db28 | 1561 | 97.3% | 1527 | 27 | 0 | 0 |
| Db29 | 1441 | 97.2% | 1527 | 28 | 0 | 0 |
| Db30 | 1501 | 97.1% | 1527 | 29 | 0 | 0 |
| Db31 | 1381 | 97.0% | 1527 | 30 | 0 | 0 |
| Db32 | 1561 | 96.9% | 1527 | 31 | 0 | 0 |
| Db33 | 1441 | 96.8% | 1527 | 32 | 0 | 0 |
| Db34 | 1501 | 96.7% | 1527 | 33 | 0 | 0 |
| Db35 | 1381 | 96.6% | 1527 | 34 | 0 | 0 |
| Db36 | 1561 | 96.5% | 1527 | 35 | 0 | 0 |
| Db37 | 1441 | 96.4% | 1527 | 36 | 0 | 0 |
| Db38 | 1501 | 96.3% | 1527 | 37 | 0 | 0 |
| Db39 | 1381 | 96.2% | 1527 | 38 | 0 | 0 |
| Db40 | 1561 | 96.1% | 1527 | 39 | 0 | 0 |
| Db41 | 1441 | 96.0% | 1527 | 40 | 0 | 0 |
| Db42 | 1501 | 95.9% | 1527 | 41 | 0 | 0 |
| Db43 | 1381 | 95.8% | 1527 | 42 | 0 | 0 |
| Db44 | 1561 | 95.7% | 1527 | 43 | 0 | 0 |
| Db45 | 1441 | 95.6% | 1527 | 44 | 0 | 0 |
| Db46 | 1501 | 95.5% | 1527 | 45 | 0 | 0 |
| Db47 | 1381 | 95.4% | 1527 | 46 | 0 | 0 |
| Db48 | 1561 | 95.3% | 1527 | 47 | 0 | 0 |
| Db49 | 1441 | 95.2% | 1527 | 48 | 0 | 0 |
| Db50 | 1501 | 95.1% | 1527 | 49 | 0 | 0 |
| Db51 | 1381 | 95.0% | 1527 | 50 | 0 | 0 |
| Db52 | 1561 | 94.9% | 1527 | 51 | 0 | 0 |
| Db53 | 1441 | 94.8% | 1527 | 52 | 0 | 0 |
| Db54 | 1501 | 94.7% | 1527 | 53 | 0 | 0 |
| Db55 | 1381 | 94.6% | 1527 | 54 | 0 | 0 |
| Db56 | 1561 | 94.5% | 1527 | 55 | 0 | 0 |
| Db57 | 1441 | 94.4% | 1527 | 56 | 0 | 0 |
| Db58 | 1501 | 94.3% | 1527 | 57 | 0 | 0 |
| Db59 | 1381 | 94.2% | 1527 | 58 | 0 | 0 |
| Db60 | 1561 | 94.1% | 1527 | 59 | 0 | 0 |
| Db61 | 1441 | 94.0% | 1527 | 60 | 0 | 0 |
| Db62 | 1501 | 93.9% | 1527 | 61 | 0 | 0 |
| Db63 | 1381 | 93.8% | 1527 | 62 | 0 | 0 |
| Db64 | 1561 | 93.7% | 1527 | 63 | 0 | 0 |
| Db65 | 1441 | 93.6% | 1527 | 64 | 0 | 0 |
| Db66 | 1501 | 93.5% | 1527 | 65 | 0 | 0 |
| Db67 | 1381 | 93.4% | 1527 | 66 | 0 | 0 |
| Db68 | 1561 | 93.3%</ | | | | |

```

Qy 112 -----ACAGTTGTTACAGTTCCTGCTCATGACGAGCTTACCCCA 150
Db 61 ACCACAGCCCCCTAAACCCCGACAGTGTGTTACGGGTTCTGTGTATGACAGCTCTAACCCCA 120
Qy 151 GGTGGAGAAAGAGAGATTGGGTTACCCAGAGAAATTCAATGCCCCAGCTCTATCTGAGAG 210
Db 121 GGTGAGAAAGAGAGATTGGGTTACCCAGAGAAATTCAATGCCCCAGCTCTATCTGAGAG 180
Qy 211 AATGCTGAGATGACAGACAGAGCTTCTCAAGCCACAGCCCCGGTTCAAGCTCTCC 270
Db 181 AATGCTGAGATGACAGAGCTTGA-----TCGAATTCGGATGTCGGGGCTCCAGC 234
Qy 271 ACCACAGAGAGAGATGCTCTGGCCCCGGGCAACGGAAACAAGCTTCAAGTTCAAGCT 330
Db 235 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 294
Qy 331 GCGACCTGGGAGAGAGATGCTCTGGCTCCAGTCCAGAGCCAGCCCTGGGCTCCAGC 390
Db 295 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 354
Qy 391 ACCCGAGCCAGCCAGATGCTCACTGAGCCCCGGAACAAGCCAGCCCGGCTCCAGC 450
Db 355 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 414
Qy 451 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGC----- 492
Db 415 GCAACCCAGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 474
Qy 493 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 552
Db 475 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 534
Qy 553 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 612
Db 535 GCGCCCCAGGCCAGGTGTCACTGGCCCCGGGACACAGGCCGGGCCCGGGCTCCAGC 594
Qy 613 CCGGCTTGGGCTCCAGCCGCTCCAGTTCACATGCTCACTGGCTTCAGGCTCTGCA 672
Db 595 CCGGCTTGGGCTCCAGCCGCTCCAGTTCACATGCTCACTGGCTTCAGGCTCTGCA 654
Qy 673 TCAGGCTCAGCTTCTGCTGTGTCACAAAGGCACTCTGCAAGGCTTACCAACCCCA 732
Db 655 TCAGGCTCAGCTTCTGCTGTGTCACAAAGGCACTCTGCAAGGCTTACCAACCCCA 714
Qy 733 GCGAGAGAGAGAGATGCTCACTGGCTCCAGCCAGCTCAATGACAGGTAACCTCTCTC 792
Db 715 GCGAGAGAGAGAGATGCTCACTGGCTCCAGCCAGCTCAATGACAGGTAACCTCTCTC 774
Qy 793 GCGAGCTAGAGAGAGATGCTCACTGGCTCCAGCCAGCTCAATGACAGGTAACCTCTCTC 852
Db 775 GCGAGCTAGAGAGAGATGCTCACTGGCTCCAGCCAGCTCAATGACAGGTAACCTCTCTC 834
Qy 853 ACCCTCTCAATCAAGAGACTTCTCCAGTGTGCTAATGGGCTCTCTTCTTTTCTCTG 912
Db 835 ACCCTCTCAATCAAGAGACTTCTCCAGTGTGCTAATGGGCTCTCTTCTTTTCTCTG 894
Qy 913 TCGTTTCAATTTAAACCTCCAGTTAATCTCTGAGAAATCCAGACCCAGATAC 972
Db 895 TCGTTTCAATTTAAACCTCCAGTTTCTCTCTCTCTGAAATCCAGACCCAGATAC 954
Qy 973 TACCAAGCTGACAGAGAGATTTCTGAAATGTTTTGCAATTTATTAACAAGGGGGT 1032
Db 955 TACCAAGCTGACAGAGAGATTTCTGAAATGTTTTGCAATTTATTAACAAGGGGGT 1014
Qy 1033 TTTTGGGCTCTCTCAATATTAAGTTCAAGCCAGAGATCTGTGTGTAACAATTGACTCTG 1092
Db 1015 TTTTGGGCTCTCTCAATATTAAGTTCAAGCCAGAGATCTGTGTGTAACAATTGACTCTG 1074
Qy 1093 GCGTTCCGAGAAAGTACATGCTCAACAGACAGTGAAGACAGTTCATCAGATTA 1152
Db 1075 GCGTTCCGAGAAAGTACATGCTCAACAGACAGTGAAGACAGTTCATCAGATTA 1134

```

```

Qy 1153 ACCGAGAGAGCTCTGATATTAACCTGACAGATCTCAGAGCTCAGGCTGAGTGTGCA 1212
Db 1135 ACCGAGAGAGCTCTGATATTAACCTGACAGATCTCAGAGCTCAGGCTGAGTGTGCA 1194
Qy 1213 TTTTCTTTCTCTGCCCAGCTGTGGGGCTGGGGCTGCCAGGCTGGGGCATTCGCCCTGCTGG 1272
Db 1195 TTTTCTTTCTCTGCCCAGCTGTGGGGCTGGGGCTGCCAGGCTGGGGCATTCGCCCTGCTGG 1254
Qy 1273 CTGTGCTGTGTTGCTGGTTGGCTGGGCAATGTCTATCTCATTTGCCCTGCTGTGAG 1332
Db 1255 CTGTGCTGTGTTGCTGGTTGGCTGGGCAATGTCTATCTCATTTGCCCTGCTGTGAG 1314
Qy 1333 TGCAGCCAGAAAGATCAAGGAGAGCTGACATCTTTCAGACCCGGGATATCCATCATCT 1392
Db 1315 TGCAGCCAGAAAGATCAAGGAGAGCTGACATCTTTCAGACCCGGGATATCCATCATCT 1374
Qy 1393 ATGAGGAGATACCCACCTACCAACCATGGGGCTATGTGCCCTTACGAGATACGAT 1452
Db 1375 ATGAGGAGATACCCACCTACCAACCATGGGGCTATGTGCCCTTACGAGATACGAT 1434
Qy 1453 CGTAGCCCTATGAGAAAGTTTCTGCAAGTATGTGGCAGAGCTCTTTACACAAC 1512
Db 1435 CGTAGCCCTATGAGAAAGTTTCTGCAAGTATGTGGCAGAGCTCTTTACACAAC 1494
Qy 1513 CCAGCAGTGGAGCCACTTCTGCAACTTGTAG 1545
Db 1495 CCAGCAGTGGAGCCACTTCTGCAACTTGTAG 1527

RESULT 8
AABD0394
ID AABD0394 standard; DNA; 1737 BP.
XX
AC AABD0394;
XX
DT 29-AUG-2000 (first entry)
XX
DB Ubiquitin-B, coli LacI-human Mucin 1 fusion protein encoding DNA #5.
XX
KW Ubiquitin; LacI; beta-galactosidase; fusion protein;
KW human; Mucin 1; MUC-1; tumour; pMR530 expression vector;
KW anti-tumour; therapy; immune response; cytosolic; vaccine; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1737
FT /tag= a
FT /product= "Ubiquitin-LacI-MUC-1 fusion protein"
FT /tag= b
FT /label= UBILacI DNA
FT /note= "includes ubiquitin-B, coli LacI fusion DNA"
FT /tag= c
FT /note= "Human MUC-1 partial DNA that corresponds to
FT nucleotides 136-1497 of the EMBL sequence J05581 with
FT two stop codons"
FT
PR W0200025827-A2.
PD 11-MAY-2000.
PE 18-OCT-1999; 99WO-BP07874.
PR 30-OCT-1998; 98IT-MI02330.
PA (MENA ) MENARINI RICERCHE SPA.
PI Parente D, Di Massimo AM, De Santis R;
XX
DR WPI; 2000-365410/31.

```


Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
metastatic liver tumour; cytostatic; expression profile; disease state;
disease progression; drug toxicity; drug efficacy; drug metabolism.
Homo sapiens.
NM00229103-A2.
11-APR-2002.
02-OCT-2001; 2001MO-US30589.
02-OCT-2000; 2000US-237054P.
(GENE-) GENE LOGIC INC.
Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
WPI; 2002-426119/45.
Diagnosing and detecting the progression of liver cancer,
hepatocellular carcinoma or metastatic liver tumor in a patient,
involves detecting the level of expression of two or more genes in a
liver tissue sample -
Claim 1; SEQ ID NO 2121; 298bp; English.
The invention relates to a novel method for diagnosing and detecting the
progression of liver cancer, hepatocellular carcinoma or metastatic liver
tumor in a patient, and differentiating metastatic liver cancer from
hepatocellular carcinoma in a patient, involving detecting the level of
expression of two or more genes represented in ABN93503-ABN97455 in a
tissue sample. The method of the invention has hepatotropic, and
cytostatic activity. The method is useful for diagnosing and detecting
the progression of liver cancer, hepatocellular carcinoma and metastatic
liver carcinoma in a patient. The method is useful for identifying
expression profiles which serve as useful diagnostic markers as well as
markers that can be used to monitor disease states, disease progression,
drug toxicity, drug efficacy and drug metabolism.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WPI
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
Best Local Similarity 94.8%; Pred. No. 2.2e-248;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 249 CAGCCCGGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 308
DB 2545 CAGCCCGGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2604
QY 309 GAAACAGCTTCAAGTTCAGTCTGACCTGGAGAGATGTCATCTGCGTCCAGTCA 368
DB 2605 CAGCCCGGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2664
QY 369 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 428
DB 2665 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2724
QY 429 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 488
DB 2725 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2784
QY 489 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 548
DB 2785 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2844
QY 549 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 608
DB 2845 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2904

QY 609 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 668
DB 2905 CAGGCGAGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 2964
QY 669 TGCATCAGGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 728
DB 2965 TGCATCAGGCTTCAAGCTCTTCACTCAGGAGAGATGTCATCTGCGCCCGCCAC 3024
QY 729 CCGAGCGAG 788
DB 3025 CCGAGCGAG 3084
QY 789 CCGAGCGAG 848
DB 3085 CCGAGCGAG 3144
QY 849 TGTACCTCTTCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 908
DB 3145 TGTACCTCTTCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3204
QY 909 CCGTCTCTTCAATCAG 968
DB 3205 CCGTCTCTTCAATCAG 3264
QY 969 CTACACCAAG 1028
DB 3265 CTACACCAAG 3324
QY 1029 GGGTTTCTGAGCTCTTCAATATTAAGTTAGAGAGAGAGAGAGAGAGAGAGAGAG 1088
DB 3325 GGGTTTCTGAGCTCTTCAATATTAAGTTAGAGAGAGAGAGAGAGAGAGAGAGAG 3384
QY 1089 TGTGAGCTTCAAG 1148
DB 3385 TGTGAGCTTCAAG 3444
QY 1149 TAAACGAG 1208
DB 3445 TAAACGAG 3504
QY 1209 GGCATTTCTTCTGAG 1268
DB 3505 GGCATTTCTTCTGAG 3564
QY 1269 GGTGAGCTTCTGAG 1328
DB 3565 GGTGAGCTTCTGAG 3624
QY 1329 TCAAGGCGAG 1388
DB 3625 TCAAGGCGAG 3684
QY 1389 TCTTATGAG 1448
DB 3685 TCTTATGAG 3744
QY 1449 CGATGTAG 1508
DB 3745 CGATGTAG 3804
QY 1509 AAACGAG 1547
DB 3805 AAACGAG 3843
RESULT 11
ID ABLE7071 standard; DNA; 4139 BP.
XX ABLE7071;
XX AC
XX DT 15-MAY-2002 (first entry)
XX

DE Thyroid cancer related gene sequence SEQ ID NO:5408.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
XX gene; ds.
OS Homo sapiens.
XX
XX WO200194629-A2.
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237122P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237288P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 02-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 5408; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical

CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL1664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms tumour.
XX
XX Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
SO
Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
Best Local Similarity 94.8%; Pred. No. 2.2e-246;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 249 CAGCCCGGATTCAAGCTCTCCACCACTCAAGGACAGAGATGTCACTTGGCCCGGACAC 308
DB 2545 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2604
QY 309 GGAACCACTTCAGATTCACTGACCACTGGGGACAGAGATGTCACTCGGATCCAGTCAAC 368
DB 2605 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2664
QY 369 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 428
DB 2665 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2724
QY 429 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 488
DB 2725 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2784
QY 489 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 548
DB 2785 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2844
QY 549 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 608
DB 2845 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2904
QY 609 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 668
DB 2905 CAGGCGCGCCCGGGCTCCACCGCCCGCCAGCCACGGTGTCACTCGGCCCGGACAC 2964
QY 669 TGCATCAGGCTCAAGCTTCTAAGTGTGACAGGACGACCTGCGAGGGCTTACACACAC 728
DB 2965 TGCATCAGGCTCAAGCTTCTAAGTGTGACAGGACGACCTGCGAGGGCTTACACACAC 3024
QY 729 CCAGGCGCAAGAGACATCCAGGATTTCCAGGACACATCTGATTAATCTCTTACAC 788
DB 3025 CCAGGCGCAAGAGACATCTCAATTCATTCACGACACATCTGATTAATCTCTTACAC 3084
QY 789 CCTTCCAGGCTTACGACCAAGATGATGACGATGACGATGACGATGACGATGACGATGAC 848
DB 3085 CCTTCCAGGCTTACGACCAAGATGATGACGATGACGATGACGATGACGATGACGATGAC 3144
QY 849 TCTCAGCTCTCCATCAACAGCACTTCTCCCAAGTGTCTACTGAGGATCTTTCTTTT 908
DB 3145 TCTCAGCTCTCCATCAACAGCACTTCTCCCAAGTGTCTACTGAGGATCTTTCTTTT 3204
QY 909 CCGTCTTTCATCTTCAATTAACCTCAATTAATCTCTCTGGAAGATCCAGACCGA 968
DB 3205 CCGTCTTTCATCTTCAATTAACCTCAATTAATCTCTCTGGAAGATCCAGACCGA 3264
QY 969 CTAATCAAGAGAGAGAGAGAGAGATTTGAAAGTTTGGAGATTTAAAGAG 1028
DB 3265 CTAATCAAGAGAGAGAGAGAGATTTGAAAGTTTGGAGATTTAAAGAG 3324


```
QY 1029 GGGTTTCTGGAGCTCTGCAATATTAAATTGAGGCGAGATCTGAGTGTGATCAATTGAC 1088
DB 3325 GGGTTTCTGGAGCTCTGCAATATTAAATTGAGGCGAGATCTGAGTGTGATCAATTGAC 3384
QY 1089 TCTGGCCCTTCGAGAGGATGATCATATGTCACGAGCTGAGAGACAGTTCAATCAGTA 1148
DB 3385 TCTGGCCCTTCGAGAGGATGATCATATGTCACGAGCTGAGAGACAGTTCAATCAGTA 3444
QY 1149 TAAACGGAAGCAGCTCTGATATATACCTGACGATCTGAGCTGAGCTGAGTCAATG 1208
DB 3445 TAAACGGAAGCAGCTCTGATATATACCTGACGATCTGAGCTGAGCTGAGTCAATG 3504
QY 1209 GCCATTTCTTTCTTCTGCGCAGTCTGGGGCTGGGGTGGCCAGGCTGAGGCTGCT 1268
DB 3505 GCCATTTCTTTCTTCTGCGCAGTCTGGGGCTGGGGTGGCCAGGCTGAGGCTGCT 3564
QY 1269 GGTGCTGATCTGTGTTCTGTTGCGCTGGCCATTTGCTATCTGATTTGCGCTGTCTG 1328
DB 3565 GGTGCTGATCTGTGTTCTGTTGCGCTGGCCATTTGCTATCTGATTTGCGCTGTCTG 3624
QY 1329 TGAAGGCGCGCGAAAGATCAAGGAGCTGAGATCTTTCCAGCCCGGAAATCCTACCA 1388
DB 3625 TGAAGGCGCGCGAAAGATCAAGGAGCTGAGATCTTTCCAGCCCGGAAATCCTACCA 3684
QY 1389 TCTATAGAGGAGTACCCCACTACACACCCATGGCGCTATGTGCCCTTACGAGTAC 1448
DB 3685 TCTATAGAGGAGTACCCCACTACACACCCATGGCGCTATGTGCCCTTACGAGTAC 3744
QY 1449 CGATCTGAGCCCTATGAGAGGTTTCTGAGGATGATGAGGAGAGGCTCTCTTACAC 1508
DB 3745 CGATCTGAGCCCTATGAGAGGTTTCTGAGGATGATGAGGAGAGGCTCTCTTACAC 3804
QY 1509 AAACCCAGCAGTGGCGAGCACTTGGCCAACTGTGATGAGG 1547
DB 3805 AAACCCAGCAGTGGCGAGCGCTTCTGCAACTGTGATGAGG 3843
DB
RESULT 12
ABL67544
ID ABL67544 standard; DNA; 4139 BP.
XX
AC ABL67544;
XX
DT 15-MAY-2002 (first entry)
XX
DE Thyroid cancer related gene sequence SEQ ID NO:5881.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancersous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN MO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001MO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209511P.
PR 18-SEP-2000; 2000US-231133P.
PR 18-SEP-2000; 2000US-231617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
```

```
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236103P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI: 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 5881; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (II)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
XX
Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
Query March 75.7% Score 1190.2; DB 24; Length 4139;
Best Local Similarity 94.8%; Pred. No.2.2e-248;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 249 CAGCCCGGTTCAAGCTCTTCCACATCAAGGACAGAGTCACTTGGCCCCGGCAC 308
DB 2545 CAGCGCGCGCCCGGCTTCACCGCCGCCCGACGGATGTCACCTTGGCGGACAC 2604
QY 309 GGAACGACTTCAGTTCACTGTCACCTGGGAGCAGAGTGTACCTGCTCCAGTAC 368
```

[illegible]

| | | | |
|-----------|------|--|------|
| Db | 3665 | TCGATGATGCGAGTACCCCACTACACACCCATG66GCGCTATGTGCCCTTACGACATAC | 3744 |
| Qy | 1449 | CGATGTGATGCCCTCTATGAGAGTTCCTTGACAGTAAATGTTGGCAGCAGCCTCTTTACAC | 1508 |
| Db | 3745 | CGATGTGATGCCCTCTATGAGAGTTCCTTGACAGTAAATGTTGGCAGCAGCAGCCTCTTTACAC | 3804 |
| Qy | 1509 | AAACCCAGCAGTGGCAGCAGCCTTCTGCCACTTGTAGG | 1547 |
| Db | 3805 | AAACCCAGCAGTGGCAGCAGCCTTCTGCCACTTGTAGG | 3843 |
| RESULT 13 | | | |
| ABL68884 | ID | ABL68884 standard; DNA; 4139 BP. | |
| XX | AC | ABL68884; | |
| XX | DT | 15-MAY-2002 (first entry) | |
| XX | DE | Kidney cancer related gene sequence SRQ ID NO.7221. | |
| XX | KW | Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid; | |
| XX | KW | stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous; | |
| KW | KW | cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma | |
| KW | gene | de. | |
| XX | OS | Homo sapiens. | |
| XX | PN | WO200194629-A2. | |
| XX | PD | 13-DEC-2001. | |
| XX | PF | 30-MAY-2001; 2001WO-US10838. | |
| XX | PR | 05-JUN-2000; 2000US-209473P. | |
| PR | PR | 05-JUN-2000; 2000US-209531P. | |
| PR | PR | 18-SEP-2000; 2000US-233133P. | |
| PR | PR | 18-SEP-2000; 2000US-233617P. | |
| PR | PR | 20-SEP-2000; 2000US-234009P. | |
| PR | PR | 20-SEP-2000; 2000US-234034P. | |
| PR | PR | 20-SEP-2000; 2000US-234052P. | |
| PR | PR | 22-SEP-2000; 2000US-234509P. | |
| PR | PR | 22-SEP-2000; 2000US-234567P. | |
| PR | PR | 22-SEP-2000; 2000US-234924P. | |
| PR | PR | 25-SEP-2000; 2000US-235082P. | |
| PR | PR | 25-SEP-2000; 2000US-235077P. | |
| PR | PR | 25-SEP-2000; 2000US-235082P. | |
| PR | PR | 25-SEP-2000; 2000US-235134P. | |
| PR | PR | 25-SEP-2000; 2000US-235280P. | |
| PR | PR | 26-SEP-2000; 2000US-235637P. | |
| PR | PR | 26-SEP-2000; 2000US-235638P. | |
| PR | PR | 27-SEP-2000; 2000US-235711P. | |
| PR | PR | 27-SEP-2000; 2000US-235720P. | |
| PR | PR | 27-SEP-2000; 2000US-235840P. | |
| PR | PR | 27-SEP-2000; 2000US-235863P. | |
| PR | PR | 28-SEP-2000; 2000US-236028P. | |
| PR | PR | 28-SEP-2000; 2000US-236032P. | |
| PR | PR | 28-SEP-2000; 2000US-236033P. | |
| PR | PR | 28-SEP-2000; 2000US-236034P. | |
| PR | PR | 28-SEP-2000; 2000US-236109P. | |
| PR | PR | 28-SEP-2000; 2000US-236111P. | |
| PR | PR | 29-SEP-2000; 2000US-236842P. | |
| PR | PR | 29-SEP-2000; 2000US-236891P. | |
| PR | PR | 02-OCT-2000; 2000US-237112P. | |
| PR | PR | 02-OCT-2000; 2000US-237113P. | |
| PR | PR | 02-OCT-2000; 2000US-237278P. | |
| PR | PR | 02-OCT-2000; 2000US-237294P. | |
| PR | PR | 02-OCT-2000; 2000US-237295P. | |
| PR | PR | 02-OCT-2000; 2000US-237316P. | |
| PR | PR | 03-OCT-2000; 2000US-237425P. | |
| PR | PR | 03-OCT-2000; 2000US-237588P. | |
| PR | PR | 03-OCT-2000; 2000US-237604P. | |
| PR | PR | 03-OCT-2000; 2000US-237606P. | |

XX 29-NOV-2001.
 PD 23-MAY-2001; 2001MO-US16895.
 XX 24-MAY-2000; 2000US-207107P.
 PR 13-JUN-2000; 2000US-211457P.
 PR 21-JUN-2000; 2000US-213673P.
 PR 03-AUG-2000; 2000US-223288P.
 PR 01-MAR-2001; 2001US-272790P.
 XX (CORI-) CORIXA CORP.
 PA Xu J, Mitcham JL, Harlocker SL, Dillon DC, Secrist H, Lodes MJ;
 PI Algate PA, Fling SP, Mannon J, Benson DR, Carter D;
 XX WPI; 2002-097641/13.
 DR New isolated polynucleotide encoding polypeptide comprising portion of
 XX PT ovarian tumour protein, useful for detection, diagnosis and therapy of
 XX human ovarian cancer -
 PS Claim 1; Page 274-275; 285pp; English.
 CC The invention relates to an isolated polynucleotide encoding a
 CC polypeptide comprising a portion of an ovarian tumour protein. The
 CC sequences of the invention are useful for stimulating an immune response
 CC and for treating ovarian cancer in a patient. An antigen presenting cell
 CC that expresses the sequences is useful for treating ovarian cancer by
 CC incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells
 CC can then be proliferated and administered to the patient to inhibit the
 CC development of cancer. The DNA sequences are useful as probes or primers
 CC for nucleic acid hybridisation to direct expression of a polypeptide in
 CC appropriate host cells. Detecting the presence of a cancer in a patient
 CC involves obtaining a biological sample from the patient, contacting the
 CC biological sample with an agent that binds to the protein, detecting the
 CC amount of protein that binds to the agent, comparing the amount of
 CC protein to a predetermined cut-off value and determining the presence of
 CC cancer. Sequences ABK09464-ABK09802 represent PCR primers and cDNA
 CC molecules encoding ovarian tumour proteins of the invention.
 XX
 SO Sequence 4139 BP; 632 A; 1910 C; 1055 G; 542 T; 0 other;
 Query Match 75.7%; Score 1190.2; DB 24; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-248;
 Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 249 CAGCCCGGATTCAGGCTCTCCACCACTCAGGAGACAGATGTCACTTGCCCCCGGACAC 308
 DB 2545 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2604
 QY 309 GGAACCGAGTTTCAGGTTCACTGTGCACTGAGGAGACAGATGTCACTTGCCCCCGGACAC 368
 DB 2605 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2664
 QY 369 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 428
 DB 2665 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2724
 QY 429 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 488
 DB 2725 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2784
 QY 489 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 548
 DB 2785 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2844
 QY 549 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 608
 DB 2845 CAGGCGGCGCCCGGGCTCCACCGCCCGCCAGCCCAAGTGTCACTTGCCCCCGGACAC 2904
 QY 609 CAGGCGGCGCTTGGGCTCCACCGCCCGCTCAGTCCCAAGTGTCACTTGCCCCCGGACAC 668
 |||||

DB 2905 CAGGCGGCGCTTGGGCTCCACCGCCCGCTCAGTCCCAAGTGTCACTTGCCCCCGGACAC 2964
 QY 669 TGCATCAGGCTCAGCTTCTTACCTGTGTGCAACAAGGACACTTGTGCAAGGCTTACCAAC 728
 DB 2965 TGCATCAGGCTCAGCTTCTTACCTGTGTGCAACAAGGACACTTGTGCAAGGCTTACCAAC 3024
 QY 729 CCAGCAGCAGAGAGCACTCCACCCAGAGATTCCACAGCAACACTGTGATCTCTTACAC 788
 DB 3025 CCAGCAGCAGAGAGCACTCCACCTCAATTCACAGCCACCACTGTGATCTCTTACAC 3084
 QY 789 CTTGCGAGCATTAGCAGCAAGACATGTGCGAGTACACTCAGTACCAAGTACCTCC 848
 DB 3085 CTTGCGAGCATTAGCAGCAAGACATGTGCGAGTACACTCAGTACCAAGTACCTCC 3144
 QY 849 TCTCACTCTCTCAATCAGACACTTCTCCAGAGTTGTACTGTGGGCTCTCTTTCTTTT 908
 DB 3145 TCTCACTCTCTCAATCAGACACTTCTCCAGAGTTGTACTGTGGGCTCTCTTTCTTTT 3204
 QY 909 CCTGCTTTTCAATTTCAACCTCCAGTTTAAATCTCTCTGAGAAAGATCCAGACCA 968
 DB 3205 CCTGCTTTTCAATTTCAACCTCCAGTTTAAATCTCTCTGAGAAAGATCCAGACCA 3264
 QY 969 CTACACCAAGACATGACAGAGACATTTCTGAAATGTTTGGAGATTTTAAACAAG 1028
 DB 3265 CTACACCAAGACATGACAGAGACATTTCTGAAATGTTTGGAGATTTTAAACAAG 3324
 QY 1029 GGGTTTTTCTGGGCTCTCCAAATTTAAGTTCAAGGCCAGATCTGTGTACAAATTGAC 1088
 DB 3325 GGGTTTTTCTGGGCTCTCCAAATTTAAGTTCAAGGCCAGATCTGTGTGTACAAATTGAC 3384
 QY 1089 TGTGGCCTTCCAGAGATGACATGATGTCAGAGCTGAGAGACAGAGTTCAATCAGTA 1148
 DB 3385 TGTGGCCTTCCAGAGATGACATGATGTCAGAGCTGAGAGACAGAGTTCAATCAGTA 3444
 QY 1149 TAAACGGAAGCAGCTCTCGATATTAACCTGACGATCTCAAGCTGACGTATGT 1208
 DB 3445 TAAACGGAAGCAGCTCTCGATATTAACCTGACGATCTCAAGCTGACGTATGT 3504
 QY 1209 GGCATTTCTTTTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGCT 1268
 DB 3505 GGCATTTCTTTTCTGCGCAGTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGCT 3564
 QY 1269 GGT 1328
 DB 3565 GGT 3624
 QY 1329 TCAAGTCCGCCGAAAGAACTACGAGGCACTGACATCTTTCAGGCTGGGATACCA 1388
 DB 3625 TCAAGTCCGCCGAAAGAACTACGAGGCACTGACATCTTTCAGGCTGGGATACCA 3684
 QY 1389 TCTATGAGGAGATACCCCACTACCAACCACTGAGGCTATGTGCTTACAGTAC 1448
 DB 3685 TCTATGAGGAGATACCCCACTACCAACCACTGAGGCTATGTGCTTACAGTAC 3744
 QY 1449 CGATCTGAGCCCTTATGAGAGAGTTTCTGCAAGTAAATGTGTGACACAGCTCTTTAC 1508
 DB 3745 CGATCTGAGCCCTTATGAGAGAGTTTCTGCAAGTAAATGTGTGACACAGCTCTTTAC 3804
 QY 1509 AAACCCAGAGTGGACCACTTCTGCAACTTGTGCAACTTGTGAGG 1547
 DB 3805 AAACCCAGAGTGGACCGCTTCTGCAACTTGTGAGG 3843
 RESULT 15
 AAQ24679
 ID AAQ24679 standard; DNA; 1074 BP.
 XX
 AC AAQ24679;
 XX
 DT 06-NOV-1992 (first entry)
 XX
 DE H23-ETa-T antigen C-terminal coding sequence.
 XX

Thu May 8 18:53:06 2003

us-09-658-621b-1.rng

Page 20

Db 1021 GTGGCAGCCACTTCTGCCACTTGTAGGGGCAAGTGGCCCTCTGAGCTGAGTGG 1074

Search completed: May 8, 2003, 15:04:24
Job time : 447 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: May 8, 2003, 14:54:17 ; Search time 2408 Seconds
(without alignments)
10572.801 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gaattccctgcgtcgttgaa.....tcgcccctgcagctcgtcgtg 1572

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estb1:*
2: em_estb2:*
3: em_estb3:*
4: em_estb4:*
5: em_estb5:*
6: em_estb6:*
7: em_estb7:*
8: em_estb8:*
9: em_estb9:*
10: em_estb10:*
11: em_estb11:*
12: em_estb12:*
13: em_estb13:*
14: em_estb14:*
15: em_estb15:*
16: em_estb16:*
17: em_estb17:*
18: em_estb18:*
19: em_estb19:*
20: em_estb20:*
21: em_estb21:*
22: em_estb22:*
23: em_estb23:*
24: em_estb24:*
25: em_estb25:*
26: em_estb26:*
27: em_estb27:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 701.8 | 44.6 | 874 | BQ918845 | BQ918845 AGNCOURT |
| 2 | 701.8 | 44.6 | 944 | BQ928821 | BQ928821 AGNCOURT |
| 3 | 659.8 | 42.0 | 682 | AL543598 | AL543598 AL543598 |
| 4 | 659.8 | 41.6 | 670 | BM817771 | BM817771 K-EST0084 |
| 5 | 642.4 | 40.9 | 798 | BG740690 | BG740690 602631536 |
| 6 | 635.8 | 40.4 | 664 | BM794696 | BM794696 K-EST0076 |

| | | | | | | |
|----|-------|------|-----|----|----------|--------------------|
| 7 | 621 | 39.5 | 670 | 12 | BF338440 | BF338440 602034094 |
| 8 | 611.4 | 38.9 | 621 | 14 | BM629853 | BM629853 K-EST0103 |
| 9 | 594.4 | 37.8 | 604 | 14 | BM850782 | BM850782 K-EST0131 |
| 10 | 589.8 | 37.5 | 593 | 14 | BM741056 | BM741056 K-EST0131 |
| 11 | 573.4 | 36.5 | 565 | 10 | BM795113 | BM795113 K-EST0076 |
| 12 | 557.2 | 35.4 | 559 | 14 | AM387919 | AM387919 MR4-ST011 |
| 13 | 555.8 | 35.4 | 559 | 14 | BM793904 | BM793904 K-EST0074 |
| 14 | 554.4 | 35.3 | 559 | 14 | BG774910 | BG774910 602649832 |
| 15 | 543.8 | 34.6 | 629 | 12 | BM791452 | BM791452 K-EST0071 |
| 16 | 541 | 34.4 | 950 | 12 | BG775831 | BG775831 602650441 |
| 17 | 539.8 | 34.3 | 927 | 12 | BG698098 | BG698098 602659979 |
| 18 | 530.8 | 33.8 | 604 | 14 | BM791359 | BM791359 K-EST0071 |
| 19 | 527.4 | 33.5 | 529 | 14 | BM633236 | BM633236 K-EST0107 |
| 20 | 525.4 | 33.4 | 524 | 14 | BM631319 | BM631319 K-EST0105 |
| 21 | 520.8 | 31.9 | 524 | 14 | BM622982 | BM622982 K-EST0093 |
| 22 | 494.6 | 31.5 | 566 | 10 | AM369405 | AM369405 QV1-BN000 |
| 23 | 487.8 | 31.0 | 580 | 9 | AI245417 | AI245417 QK29C06.X |
| 24 | 486.8 | 31.0 | 502 | 14 | BM746832 | BM746832 K-EST0021 |
| 25 | 478.4 | 30.4 | 480 | 14 | BM652127 | BM652127 K-EST0133 |
| 26 | 474 | 30.2 | 722 | 9 | AL543556 | AL543556 AL543556 |
| 27 | 466.2 | 29.7 | 471 | 14 | BM745435 | BM745435 K-EST0019 |
| 28 | 459.8 | 29.2 | 759 | 14 | BM682837 | BM682837 UI-CF-EM1 |
| 29 | 454.6 | 28.9 | 770 | 14 | BM691344 | BM691344 UI-H-D10 |
| 30 | 447.8 | 28.5 | 725 | 13 | BI761034 | BI761034 603048570 |
| 31 | 447.6 | 28.5 | 472 | 14 | BM759495 | BM759495 K-EST0039 |
| 32 | 447.6 | 28.5 | 485 | 13 | BI014455 | BI014455 PM0-ET020 |
| 33 | 444.6 | 28.3 | 760 | 14 | BM693267 | BM693267 UI-H-D10 |
| 34 | 427.4 | 27.3 | 437 | 14 | BM686872 | BM686872 UI-E-CR1 |
| 35 | 425.4 | 27.1 | 438 | 14 | BM620231 | BM620231 K-EST0088 |
| 36 | 424.4 | 26.9 | 424 | 14 | BM797973 | BM797973 K-EST0081 |
| 37 | 419.2 | 26.7 | 590 | 9 | AA429320 | AA429320 ZW07d11.X |
| 38 | 419.2 | 26.7 | 719 | 14 | BM684687 | BM684687 UI-CF-EC1 |
| 39 | 414 | 26.3 | 442 | 10 | AM579047 | AM579047 RCO-C7037 |
| 40 | 408.2 | 26.0 | 745 | 10 | AM001137 | AM001137 M24905.X |
| 41 | 407.8 | 25.9 | 711 | 14 | BM674802 | BM674802 UI-CF-EC1 |
| 42 | 405.8 | 25.8 | 707 | 14 | BQ446679 | BQ446679 UI-H-EU1 |
| 43 | 403.4 | 25.7 | 908 | 9 | AA826345 | AA826345 of11h02.5 |
| 44 | 398.4 | 25.3 | 630 | 9 | AI905920 | AI905920 PM-ET103 |
| 45 | 398.2 | 25.3 | 485 | 10 | AM369421 | AM369421 QV1-BN000 |

ALIGNMENTS

RESULT 1
LOCUS BQ918845 874 bp mRNA linear EST 20-AUG-2002
DEFINITION AGNCOURT 877143 NIH_MGC_40 Homo sapiens CDNA clone IMAGE:6384674
5', mRNA Sequence.
ACCESSION BQ918845
VERSION BQ918845.1 GI:22333543
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: ILCLM2580 row: b column: 03
High quality sequence stop: 621.
Location/Qualifiers
1..874

FEATURES
source


```
Db 363 CAGTCTGGGGGCTGGGGGTCAGAGCTGGGGGCAATCCGCTGCTGGTGTGTCTGTCT 422
Qy 1287 GATTGGCTGGGCAATTTGTCTATCTCATTTGCTGTGTGTGTGTGTGTGTGTGTGTGT 1346
Db 423 GATTGGCTGGGCAATTTGTCTATCTCATTTGCTGTGTGTGTGTGTGTGTGTGTGTGT 482
Qy 1347 CTAGGGGAGCTGGGAGCATCTTCCAGCCGGAGATACATCATATGAGAGAGATACCC 1406
Db 483 CTAGGGGAGCTGGGAGCATCTTCCAGCCGGAGATACATCATATGAGAGATACCC 542
Qy 1407 CACTTACCAACCCATGGGGGCTATGTGCTCCCTTACGAGTACCATGTATAGCCCTTAC 1466
Db 543 CACTTACCAACCCATGGGGGCTATGTGCTCCCTTACGAGTACCATGTATAGCCCTTAC 602
Qy 1467 GAAGGTTTCTGAGAGTAAATGTGTGAGAGAGCTCTCTTACCAAAACCCAGAGTGGAGC 1526
Db 603 GAAGGTTTCTGAGAGTAAATGTGTGAGAGAGCTCTCTTACCAAAACCCAGAGTGGAGC 662
Qy 1527 CACTTGGCAACTTGTATGGGGGAGCTGGGCTGTGAGCTGAGTGG 1572
Db 663 CACTTGGCAACTTGTATGGGGGAGCTGGGCTGTGAGCTGAGTGG 708

RESULT 3
LOCUS AL543598 682 bp mRNA linear EST 16-FEB-2001
DEFINITION AL543598 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1006Y07 5
prime, mRNA sequence.
ACCESSION AL543598
VERSION AL543598.1 GI:12876077
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 682)
AUTHORS Li W.B., Gruber C., Jesse J. and Polayres D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
    source
        1..682
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0D1006Y07"
            /clone_id="LTI_NFL006_P12"
            /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Life Technologies. Contact : Feng Liang Life Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT 150 a 221 c 144 g 167 t
ORIGIN
```

```
Db 61 TCGGCTCAGAGCTCTGATCAGGCTCAGCTTCTACTTGTGTGCAAAAGGACCTTGGC 120
Qy 715 AGGCTACCAACACCCAGCAGCAGAGACATCCACCCAGATTTCCAGCAGCATCTG 774
Db 121 AGGCTACCAACACCCAGCAGCAGAGACATCCACCCAGATTTCCAGCAGCATCTG 180
Qy 775 GATTCCTCTACACCCCTTGGCAGGATATGACCAAGATGATGAGATGATGATGATGAT 834
Db 181 GATTCCTCTACACCCCTTGGCAGGATATGACCAAGATGATGATGATGATGATGAT 240
Qy 835 AGCAGCTACCTCTCTTACCTCTGCAATCAGAGATCTTCCAGGTTGTACTGTGG 894
Db 241 AGCAGGATCTCTCTTACCTCTGCAATCAGAGATCTTCCAGGTTGTACTGTGG 300
Qy 895 GTCTCTTCTTCTTCTCTGCTTCTTCAATTTCAAACTTCAGATTAAATCTCTGTGAA 954
Db 301 GTCTCTTCTTCTTCTCTGCTTCTTCAATTTCAAACTTCAGATTAAATCTCTGTGAA 360
Qy 955 GATCCAGCAGCAGACTACTACCAAGAGCTGAGAGACATTTCTGAATGTTTTCAG 1014
Db 361 GATCCAGCAGCAGACTACTACCAAGAGCTGAGAGACATTTCTGAATGTTTTCAG 420
Qy 1015 ATTATTAACAAGGGGGTTTCTGGGCTCTCCATATTAAGTTCAAGGAGATCTGG 1074
Db 421 ATTATTAACAAGGGGGTTTCTGGGCTCTCCATATTAAGTTCAAGGAGATCTGG 480
Qy 1075 GTGTACATTTGACTGTGGCTTCCGAAAGTACATCATATGTCACAGCTGGAGACA 1134
Db 481 GTGTACATTTGACTGTGGCTTCCGAAAGTACATCATATGTCACAGCTGGAGACA 539
Qy 1135 CAGTTCATCATGATTAATAACGAGAGAGCTCTCATATTAACGATCTGAGAGCTG 1194
Db 540 CAGTTCATCATGATTAATAACGAGAGAGCTCTCATATTAACGATCTGAGAGCTG 599
Qy 1195 ACCGGAGTATGATGATGATTTCTTCTCTGAGGAGTGGGGTGGAGAGCTGG 1254
Db 600 ACCGGAGTATGATGATGATTTCTTCTCTGAGGAGTGGGGTGGAGAGCTGG 659
Qy 1255 GGCAATCCGCGCTGCTGGTGGT 1277
Db 660 GGCAATCCGCGCTGCTGGTGGT 682

RESULT 4
LOCUS BM817771 670 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0084243 S21SNUS2081 Homo sapiens cDNA clone S21SNUS2081-2-D07
5', mRNA sequence.
ACCESSION BM817771
VERSION BM817771.1 GI:19174184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 670)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boseun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.krdb.re.kr
Plote: 2 row: D column: 07
High quality sequence stop: 670.
Location/Qualifiers
    1..670
        /organism="Homo sapiens"
```

```

/db_xref="taxon:9606"
/clone="S21SNU52081-2-D07"
/clone_lib="S21SNU52081"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pTZ19RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A) RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deacapped
with tabacco acid pyrophosphatase (TAP). The deacapped
inact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed mRNA was
adjusted to have about 60nt. The cDNA vector was
clonitized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and Nid14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."
BASE COUNT      146 a      188 c      170 g      166 t
ORIGIN
Query Match      41.6%; Score 653.8; DB 14; Length 670;
Best Local Similarity 99.7%; Pred. No. 4.5e-136;
Matches 655; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 916 TTTCACATTGAACTCCAGTTTATCTCTCTGGAAGATCCGACGAGCTACTAC 975
Db 1 TTTCACATTGAACTCCAGTTTATCTCTCTGGAAGATCCGACGAGCTACTAC 60
Qy 976 CAAAGAGCTCCAGAGACATTTCTGAAATGTTTTCAGATTTTAAACAAGGGGTTT 1035
Db 61 CAAAGAGCTCCAGAGACATTTCTGAAATGTTTTCAGATTTTAAACAAGGGGTTT 120
Qy 1036 CTGGGCTCTCCATATTAAGTTCAAGCCAGAGATCTGTGTGATCAATTAATCTG 1095
Db 121 CTGGGCTCTCCATATTAAGTTCAAGCCAGAGATCTGTGTGATCAATTAATCTG 180
Qy 1096 TTCCGAGAAAGTACATCATATGTCACGACGTGAGACACAGTTCAATAGTAAACG 1155
Db 181 TTCCGAGAAAGTACATCATATGTCACGACGTGAGACACAGTTCAATAGTAAACG 240
Qy 1156 GAAGAGAGCTCCAGATTAACCTGAGAGATCTGAGAGCTGAGTGTGATTCGATTT 1215
Db 241 GAAGAGAGCTCCAGATTAACCTGAGAGATCTGAGAGCTGAGTGTGATTCGATTT 300
Qy 1216 CTTTCTTCGCGAGTCTGGGGCTGGGGCTGAGGCTGGGGCATCGCGTGTGCTG 1275
Db 301 CTTTCTTCGCGAGTCTGGGGCTGGGGCTGAGGCTGGGGCATCGCGTGTGCTG 360
Qy 1276 GTCTGTGTTCTGGTTGGCTGGCCATTTGTAATCTCAATGCTTGGCTGTCTG 1335
Db 361 GTCTGTGTTCTGGTTGGCTGGCCATTTGTAATCTCAATGCTTGGCTGTCTG 420
Qy 1336 CGCGGAAAGAACTAGCGAGCTGACATCTTCCAGCCGGGATCACTAATCACTATG 1395

```

```

Db 421 CGCGGAAAGAACTAGCGAGCTGAGACATCTTTCAGCCCGGATACCTACCTTATG 480
Qy 1396 AGCGAGTACCCCACTTACACACCCATGCGCTATGTCGCCCTTACGAGTACCATCT 1455
Db 481 AGCGAGTACCCCACTTACACACCCATGCGCTATGTCGCCCTTACGAGTACCATCT 540
Qy 1456 AGCCCTTATGAGAGGTTTCTGACAGTAAATGTTGGACGAGGCTCTTTCACAAACCA 1515
Db 541 AGCCCTTATGAGAGGTTTCTGACAGTAAATGTTGGACGAGGCTCTTTCACAAACCA 600
Qy 1516 GCAGTGGAGGACCTTGTGCAACTTTGAGGGGACGTCGCCCTTGTAGCTGAGTGG 1572
Db 601 GCAGTGGAGGACCTTGTGCAACTTTGAGGGGACGTCGCCCTTGTAGCTGAGTGG 657

RESULT 5
Bg740690      798 bp      mRNA      linear      EST 15-MAY-2001
LOCUS         602631536F1 NCI_CGAP_Skn3 Homo sapiens cDNA IMAGE:477632 5',
DEFINITION    mRNA sequence.
ACCESSION     Bg740690
VERSION       Bg740690.1 GI:14051343
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 798)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgabs-remail.nih.gov
               Tissue Procurement: James Cleaver, M.D.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA
               Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNLN at:
               http://image.llnl.gov
               Plate: LHAM10629 row: h column: 09
               High quality sequence stop: 798.
               Location/Qualifiers
               source
               1. 798
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:477632"
               /clone_lib="NCI CGAP_Skn3"
               /lab_host="DH10B (T1 phage-resistant)"
               /note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI;
               Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.
               Average insert size 1.5kb. Library constructed by Life
               Technologies. Note: this is a NCI CGAP Library."
BASE COUNT    177 a      224 c      198 g      199 t
ORIGIN
Query Match      40.9%; Score 642.4; DB 12; Length 798;
Best Local Similarity 99.7%; Pred. No. 1.7e-133;
Matches 654; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 817 GCCAGTAGACTCACTAGACAGGATCTCTGACCTCCCAATCAAGACATCTTCT 876
Db 1 GCCAGTAGACTCACTAGACAGGATCTCTGACCTCCCAATCAAGACATCTTCT 60
Qy 877 CCCAGTTTCTACTGGGGCTCTTCTTTCTGCTCTTTCACATTTGAAACCTCCG 936
Db 61 CCCAGTTTCTACTGGGGCTCTTCTTTTCTGCTCTTTCACATTTGAAACCTCCG 120
Qy 937 TTTAATCTCTCTGGAAGATCCAGACCG-ACCTACTACAAAGCTGCGAGAGCAT 995
Db 121 TTTAATCTCTCTGGAAGATCCAGACCGACCTACTACAAAGCTGCGAGAGCAT 180

```

```

Qy 996 TTCTGAAATGTTTTGAGATTTATTAACAAGGGGTTTCTGGGCTCTCCAAATATTA 1055
Db 181 TTCTGAAATGTTTTGAGATTTATTAACAAGGGGTTTCTGGGCTCTCCAAATATTA 240
Qy 1056 GTTCAGGACGAGATGTGGGTGATACATTTGACTGGCTTCGGAAGAGTACATCA 1115
Db 241 GTTCAGGACGAGATGTGGGTGATACATTTGACTGGCTTCGGAAGAGTACATCA 300
Qy 1116 TGTCCACGACGTCGAGACACAGTTCAATCAATATTAACGAGAGCTCTCCATATA 1175
Db 301 TGTCCACGACGTCGAGACACAGTTCAATCAATATTAACGAGAGCTCTCCATATA 360
Qy 1176 CCGAAGATCTCAGACGTCAGCGTGAATGATGTCATTTCTTTCTGCGCCAGTCTGG 1235
Db 361 CCGAAGATCTCAGACGTCAGCGTGAATGATGTCATTTCTTTCTGCGCCAGTCTGG 420
Qy 1236 GGCCTGGGGTCCAGGCTGGGGCATCCGGGCTGGTGTGTTCTGGTTGGGCT 1295
Db 421 GGCCTGGGGTCCAGGCTGGGGCATCCGGGCTGGTGTGTTCTGGTTGGGCT 480
Qy 1296 GGCATTTGTTATCTCATATGCTTGGCTGTCTGTCAAGTCCGGGAAAGACTACGGGCA 1355
Db 481 GGCATTTGTTATCTCATATGCTTGGCTGTCTGTCAAGTCCGGGAAAGACTACGGGCA 540
Qy 1356 GCTGACATCTTTCAGCCCGGAGATCTACCATCTATAGCGAGTACCCCACTACCA 1415
Db 541 GCTGACATCTTTCAGCCCGGAGATCTACCATCTATAGCGAGTACCCCACTACCA 600
Qy 1416 CACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTGGCCCTATAGAGAG 1471
Db 601 CACCCATGGGCGCTATGTGCCCCCTAGCAGTACCGATCGTGGCCCTATAGAGAG 656

```

RESULT 6

```

LOCUS BM794696 664 bp mRNA linear EST 05-MAR-2002
DEFINITION K-EST0076148 S22SNUL6n1 Homo sapiens cDNA clone S22SNUL6n1-48-A12
5', mRNA sequence.
ACCESSION BM794696
VERSION BM794696.1 GI:19142928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.U., Cheong, J.B., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 48 row: A column: 12
High quality sequence stop: 664.

```

FEATURES

```

source
1..664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNUL6n1-48-A12"
/clone_id="S22SNUL6n1"
/sex="F"
/tissue_type="Aesthetes"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pUT73-Pac; Site 1: EcorI;
Site 2: NotI; The S22SNUL6 library was contributed by the

```

```

BASE COUNT 145 a 185 c 170 g 164 t
ORIGIN
Query Match 40.4%; Score 635.8; DB 14; Length 664;
Best Local Similarity 99.5%; Pred. No. 4.9e-132;
Matches 648; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

```

Qy 922 ATTTCAACTCCAGTTTATTTCTCTGGAAGATCCAGGACCGACTACCAAGAG 981
Db 1 ATTTCAACTCCAGTTTATTTCTCTGGAAGATCCAGGACCGACTACCAAGAG 60
Qy 982 CTGACAGAGACATTTCTGGAATGTTTTGACATTTATTAACAAGGGGTTTCTGGGCT 1041
Db 61 CTGACAGAGACATTTCTGGAATGTTTTG-AGATTTATTAACAAGGGGTTTCTGGGCT 119
Qy 1042 CTCTCAATTTAAGTTCAAGGCGAGATCTGTGTGTATCAATTTGACTTGGCTTCCCA 1101
Db 120 CTCTCAATTTAAGTTCAAGGCGAGATCTGTGTGTATCAATTTGACTTGGCTTCCCA 179
Qy 1102 GAAAGTACATCAATGTCACAGCGTGAAGACAGTTCAATCAATGATTAACGGAAGA 1161
Db 180 GAAAGTACATCAATGTCACAGCGTGAAGACAGTTCAATCAATGATTAACGGAAGA 239
Qy 1162 GCCTCTGATATACCTGACAGATCTCAGAGAGTCAAGTCAATGATGATGATGATGATG 1221
Db 240 GCCTCTGATATACCTGACAGATCTCAGAGAGTCAAGTCAATGATGATGATGATGATG 299
Qy 1222 TCTGCGCAATCTGGGCTGGGGTGGCGAGGCGTGGGATCGGCTGTGGTGGTGGTGGT 1281
Db 300 TCTGCGCAATCTGGGCTGGGGTGGCGAGGCGTGGGATCGGCTGTGGTGGTGGTGGT 359
Qy 1282 GTTCTGTGGCTGGGCTGGGCTGTGTATCTCAATGCTTGGCTGTGTGTGTGTGTGTGT 1341
Db 360 GTTCTGTGGCTGGGCTGGGCTGTGTATCTCAATGCTTGGCTGTGTGTGTGTGTGTGT 419
Qy 1342 AAGAACTACGAGGAGCTGAGCATCTTTCAGCCCGGAGATACCTACATCTATAGAGAG 1401
Db 420 AAGAACTACGAGGAGCTGAGCATCTTTCAGCCCGGAGATACCTACATCTATAGAGAG 479
Qy 1402 TACCCCACTACACACACACACACACACACACACACACACACACACACACACACAC 1461
Db 480 TACCCCACTACACACACACACACACACACACACACACACACACACACACACACAC 539
Qy 1462 TATGAGAGGTTTCTGAGTATATGATGAGCAGCTCTCTTACACAAACCGAGAGTG 1521
Db 540 TATGAGAGGTTTCTGAGTATATGATGAGCAGCTCTCTTACACAAACCGAGAGTG 599
Qy 1522 GCAGCACTTCTGCAACTTGTAGGGGAGCGTCCCTGTGAGCTGAGTGG 1572
Db 600 GCAGCACTTCTGCAACTTGTAGGGGAGCGTCCCTGTGAGCTGAGTGG 650

```

RESULT 7

```

LOCUS BF338440 670 bp mRNA linear EST 22-NOV-2000
DEFINITION 602034094F2 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4162286
5', mRNA sequence.
ACCESSION BF338440
VERSION BF338440.1 GI:11284843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 670)
NIH-MGC http://mgs.nci.nih.gov/.

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcgabs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM9496 row: k column: 23
High quality sequence stop: 669.

FEATURES
source
1. 670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:418286"
/clone_1b="NCI CGAP Bm64"
/tissue_type="glio_blastoma with EGFR amplification"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP library."
Location/Qualifiers
BASE COUNT 151 a 221 c 136 g 162 t

Query Match 39.5%; Score 621; DB 12; Length 670;
Best Local Similarity 98.3%; Pred. No. 1e-128;
Matches 638; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 598 GCGCCGAGCAACAGGCGCGCTTGGGCTCCAGCGCCCTCCAGTCCCAATGTCACCTGG 657
Db 22 GTCCGGGCAACAGGCGCGCTTGGGCTCCAGCGCCCTCCAGTCCCAATGTCACCTGG 61
QY 658 GCGCTAGGCTCTGATAGGCTGAGCTTACTCTGTGTGCAACAAGGCACTTCTCCAGG 717
Db 82 GCGCTAGGCTCTGATAGGCTGAGCTTACTCTGTGTGCAACAAGGCACTTCTCCAGG 141
QY 718 GGTACCAACACCCAGCAGCAAGAGCACTCCACCGAGCACTCCAGCAGCACTGTGAT 777
Db 142 GGTACCAACACCCAGCAGCAAGAGCACTCCATTCATTCACAGCAGCACTGTGAT 201
QY 778 ACTCTACCACTCTGCGAGCCATAGCAACAAGAGTATGCGATGACATCACCATTAGC 837
Db 202 ACTCTACCACTCTGCGAGCCATAGCAACAAGAGTATGCGATGACATCACCATTAGC 261
QY 838 AGGTAACCTCTCACTCTCCCAATCAAGCACTTCCCAAGTGTCTACTAGGGGTC 897
Db 262 AGGTAACCTCTCACTCTCCCAATCAAGCACTTCCCAAGTGTCTACTAGGGGTC 321
QY 898 TCTTTCTTTTCTGCTGCTTTTCAATTTCAATTTCAATTTCTCTGTGGAAGAT 957
Db 322 TCTTTCTTTTCTGCTGCTTTTCAATTTCAATTTCAATTTCTCTGTGGAAGAT 381
QY 958 CCCAGACCGAGCTACTACCAAGAGCTGAGAGAGCACTTCTGAATGTTTTTTCAGATT 1017
Db 382 CCCAGACCGAGCTACTACCAAGAGCTGAGAGAGCACTTCTGAATGTTTTTTCAGATT 441
QY 1018 TTTTAAACAAGGGGTTTTTCTGGGCTCTCCAAATTTAAAGTTAGGCGCAGAGATGTGGT 1077
Db 442 TTTTAAACAAGGGGTTTTTCTGGGCTCTCCAAATTTAAAGTTAGGCGCAGAGATGTGGT 501
QY 1078 GTTAAATGATCTGCGGCTTCCGAAAGGATCAATGATCTCAAGAGCTGAGAGCAAG 1137
Db 502 GTTAAATGATCTGCGGCTTCCGAAAGGATCAATGATCTCAAGAGCTGAGAGCAAG 561
QY 1138 TTCAATCAATTAACAAGGAGGCTTCTGATTTAACTGATGATCTCAAGATCTCAAGCTGAGC 1197
Db 562 TTCAATCAATTAACAAGGAGGCTTCTGATTTAACTGATTTAACTGATGATCTCAAGCTGAGC 620
QY 1198 GTGAGTCAATGTCAGATTCTCTCTGCGCAATGTCGGGCTGGGGTGC 1246

Db 621 GTGAGTCAATGTCAGATTCTCTCTGCGCAATGTCGGGCTGGGGTGC 669

RESULT 8
LOCUS BM629853 621 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0103028 S21SNUS2061 Homo sapiens cDNA clone S21SNUS2061-16-E03
5', mRNA sequence.
ACCESSION BM629853
VERSION BM629853.1 GI:19186262
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 621)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim JS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kr.ibm.re.kr
Plate: 16 row: E column: 03
High quality sequence stop: 621.
Location/Qualifiers

FEATURES
source
1. 621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNUS2061-16-E03"
/clone_1b="S21SNUS2061"
/sex="F"
/tissue_type="Stomach"
/cell_type="F104ting aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pTZ19BP1; Site_1: EcoRI;
Site_2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then dephosphorylated
with tobacco acid pyrophosphatase (TAP). The dephosphorylated
intact mRNA was ligated with DNA-RNA linker including EcoRI
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
electroporated with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dG)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNA prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformation of the
remaining DNA into competent cells E. coli Top10F with
electroporation method."

BASE COUNT 137 a 188 c 134 g 162 t

Query Match 38.9%; Score 611.4; DB 14; Length 621;
 Best Local Similarity 99.0%; Pred. No. 1,4e-126;
 Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy 676 GGCTGAGCTTCTACTCTGTGTGCAACAGGCACTTGTGCAAGGGCTACCAACCCCAAGC 735
    |||||
Db 1 GGCTGAGCTTCTACTCTGTGTGCAACAGGCACTTGTGCAAGGGCTACCAACCCCAAGC 60
Qy 736 AGCAAGAGACATCCACCCAGCATTTCCAGCAGCACTCTGTACTCTCTACCACTTGGC 795
    |||||
Db 61 AGCAAGAGACATCCATTTCTCAATTTCCAGCAGCACTCTGTACTCTCTACCACTTGGC 120
Qy 796 AGCCATAGACCAAGACTGATGCGATGACATCCATAGCAGCGTACTCTCTTCCACC 855
    |||||
Db 121 AGCCATAGACCAAGACTGATGCGATGACATCCATAGCAGCGTACTCTCTTCCACC 180
Qy 856 TCCTCCATCAAGACACTTCTCCCGAGTTGTCTACTGGGGTCTTCTTTTCTGTCT 915
    |||||
Db 181 TCCTCCATCAAGACACTTCTCCCGAGTTGTCTACTGGGGTCTTCTTTTCTGTCT 240
Qy 916 TTTCACATTTCAAACCTCCAGTTTAATTCCTCTGTGAAGTCCAGCAGCACTACTAC 975
    |||||
Db 241 TTTCACATTTCAAACCTCCAGTTTAATTCCTCTGTGAAGTCCAGCAGCACTACTAC 300
Qy 976 CAAAGCTGCAAGAGACATTTCTGAAATGTTTGTGCAATTAATAAAGGGGTTT 1035
    |||||
Db 301 CAAAGCTGCAAGAGACATTTCTGAAATGTTTGTGCAATTAATAAAGGGGTTT 360
Qy 1036 CTGGGCTCTCCAAATTAAGTTCAAGCCAGGATGTGTGTGAACAATTGACTGTGCC 1095
    |||||
Db 361 CTGGGCTCTCCAAATTAAGTTCAAGCCAGGATGTGTGTGAACAATTGACTGTGCC 420
Qy 1096 TTCCGAGAGAGTACCATCAATGTCCAGACGTGAGAGACACAGTTCAATCATGATAAAG 1155
    |||||
Db 421 TTCCGAGAGAGTACCATCAATGTCCAGACGTGAGAGACACAGTTCAATCATGATAAAG 480
Qy 1156 GAGCAGAGCTCTCCATTAATACCTGAGATCTGACAGCTGACGCGTGAATGATGTGCATT 1215
    |||||
Db 481 GAGCAGAGCTCTCCATTAATACCTGAGATCTGACAGCTGACGCGTGAATGATGTGCATT 540
Qy 1216 CTTTCTCTGCGCCAGCTGTGGGGCTGGGGTGCAGGCTGGGGCATCGCGCTGTGCTG 1275
    |||||
Db 541 CTTTCTCTGCGCCAGCTGTGGGGCTGGGGTGCAGGCTGGGGCATCGCGCTGTGCTG 600
Qy 1276 GTCTGTGTCTGTGTGCGCTG 1296
    |||||
Db 601 GTCTGTGTCTGTGTGCGCTG 621
  
```

RESULT 9
 LOCUS BM850782 604 bp mRNA linear EST 06-MAR-2002
 DEFINITION K-EST0131423 S21SNUS20 Homo sapiens CDNA clone S21SNUS20-49-F07 5',
 mRNA sequence.
 ACCESSION BM850782
 VERSION BM850782.1 GI:19207181
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 604)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470

Fax: +82-42-860-4409
 Email: yongsung@mail.krrib.re.kr
 Plate: 49 row: F column: 07
 High quality sequence stop: 604.
 Location/Qualifiers

FEATURES

source

```

1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNUS20-49-F07"
/clone_1id="S21SNUS20"
/sex="F"
/tissue_type="Stomach"
/cell_type="floating aggregates"
/lab_host="Top10F"
/notes="Organ: Stomach; Vector: pTZ19RP1; Site: 1; EcoRI;
Site: 2; NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okazaki-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
  
```

BASE COUNT

137 a 185 c 127 g 155 t

ORIGIN

Query Match 37.8%; Score 594.4; DB 14; Length 604;
 Best Local Similarity 99.0%; Pred. No. 9e-123;
 Matches 598; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy 676 GGCTGAGCTTCTACTCTGTGTGCAACAGGCACTTGTGCAAGGGCTACCAACCCCAAGC 735
    |||||
Db 1 GGCTGAGCTTCTACTCTGTGTGCAACAGGCACTTGTGCAAGGGCTACCAACCCCAAGC 60
Qy 736 AGCAAGAGACATCCACCCAGCATTTCCAGCAGCACTCTGTACTCTCTACCACTTGGC 795
    |||||
Db 61 AGCAAGAGACATCCATTTCTCAATTTCCAGCAGCACTCTGTACTCTCTACCACTTGGC 120
Qy 796 AGCCATAGACCAAGACTGATGCGATGACATCCATAGCAGCGTACTCTCTTCCACC 855
    |||||
Db 121 AGCCATAGACCAAGACTGATGCGATGACATCCATAGCAGCGTACTCTCTTCCACC 180
Qy 856 TCCTCCATCAAGACACTTCTCCCGAGTTGTCTACTGGGGTCTTCTTTTCTGTCT 915
    |||||
Db 181 TCCTCCATCAAGACACTTCTCCCGAGTTGTCTACTGGGGTCTTCTTTTCTGTCT 240
Qy 916 TTTCACATTTCAAACCTCCAGTTTAATTCCTCTGTGAAGTCCAGCAGCACTACTAC 975
    |||||
Db 241 TTTCACATTTCAAACCTCCAGTTTAATTCCTCTGTGAAGTCCAGCAGCACTACTAC 300
Qy 976 CAAAGCTGCAAGAGACATTTCTGAAATGTTTGTGCAATTAATAAAGGGGTTT 1035
    |||||
Db 301 CAAAGCTGCAAGAGACATTTCTGAAATGTTTGTGCAATTAATAAAGGGGTTT 360
Qy 1036 CTGGGCTCTCCAAATTAAGTTCAAGCCAGGATGTGTGTGAACAATTGACTGTGCC 1095
    |||||
Db 361 CTGGGCTCTCCAAATTAAGTTCAAGCCAGGATGTGTGTGAACAATTGACTGTGCC 420
Qy 1096 TTCCGAGAGAGTACCATCAATGTCCAGACGTGAGAGACACAGTTCAATCATGATAAAG 1155
    |||||
Db 421 TTCCGAGAGAGTACCATCAATGTCCAGACGTGAGAGACACAGTTCAATCATGATAAAG 480
Qy 1156 GAGCAGAGCTCTCCATTAATACCTGAGATCTGACAGCTGACGCGTGAATGATGTGCATT 1215
    |||||
Db 481 GAGCAGAGCTCTCCATTAATACCTGAGATCTGACAGCTGACGCGTGAATGATGTGCATT 540
  
```

QY 1216 CCTTCTGCGCCAGTCTGGGGCTGGAGGCTGGAGGATCGGCTGCTGCTG 1275
 |||||
 Db 541 CTTTCTCTGCGCCAGTCTGGGGCTGGAGGCTGGAGGATCGGCTGCTGCTG 600
 |||||
 QY 1276 GTCCT 1279
 |||||
 Db 601 GTCCT 604

RESULT 10
 BM741056 593 bp mRNA linear EST 01-MAR-2002
 LOCUS K-EST0013425 S6SNU620 Homo sapiens cDNA clone S6SNU620-2-G04 5',
 DEFINITION mRNA sequence.
 ACCESSION BM741056
 VERSION BM741056
 KEYWORDS
 SOURCE human,
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 593)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kr.ibm.re.kr
 Plate: 27 row: G column: 04
 High quality sequence stop: 593.
 Location/Qualifiers

FEATURES
 source

1..593
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S6SNU620-2-G04"
 /clone_1ib="S6SNU620"
 /sex="F"
 /tissue_type="Asclies"
 /cell_type="Scattering floating"
 /cell_line="SNU-620"
 /lab_host="Top10P"
 /note="organ: Stomach; Vector: pCMS; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tabacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dt-selected mRNA by
 priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10P by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."
 BASE COUNT 127 a 161 c 159 g 146 t
 ORIGIN

Query Match 37.5%; Score 589.8; DB 14; Length 593;
 Best Local Similarity 99.7%; Pzed. No. 9,6e-122;
 Matches 591; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 977 AAGAGCTCAGAGAGACATTCTGAATGTTTGCAGATTATAACAAGGGGTTTTC 1036
 |||||

Db 1 AAGAGCTCAGAGAGACATTCTGAATGTTTGCAGATTATAACAAGGGGTTTTC 60
 QY 1037 TGGGCTCTCCCAATTATTAAGTTCAGGCGAGATCTGTGTGTAACAATTGACTTGGCT 1096
 |||||
 Db 61 TGGGCTCTCCCAATTATTAAGTTCAGGCGAGATCTGTGTGTAACAATTGACTTGGCT 120
 |||||
 QY 1097 TCCGAGAGGTAACATCAATGTCACACAGCTGGAGACACAGTTCAATCATATAAAGG 1156
 |||||
 Db 121 TCCGAGAGGTAACATCAATGTCACACAGCTGGAGACACAGTTCAATCATATAAAGG 180
 |||||
 QY 1157 AAGAGCTCTCGAATAAATCACTGACGATCTCAGAGCTGACGCTGATGTCATTTC 1216
 |||||
 Db 181 AAGAGCTCTCGAATAAATCACTGACGATCTCAGAGCTGACGCTGATGTCATTTC 240
 |||||
 QY 1217 CTTTCTCTGCGCCAGTCTGGGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 1276
 |||||
 Db 241 CTTTCTCTGCGCCAGTCTGGGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 300
 |||||
 QY 1277 TCTGTGTCGTGGTGGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTG 1336
 |||||
 Db 301 TCTGTGTCGTGGTGGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTG 360
 |||||
 QY 1337 GCCGAGAGACTACGCGCAGCTGACATCTTTCCAGCCCGGAGTACCTACATCTATGA 1396
 |||||
 Db 361 GCCGAGAGACTACGCGCAGCTGACATCTTTCCAGCCCGGAGTACCTACATCTATGA 420
 |||||
 QY 1397 GCGAGTACCCACCTACACACACCCATGGGCGCTATGTGCCCTAGCAGTACCGATCGTA 1456
 |||||
 Db 421 GCGAGTACCCACCTACACACCCATGGGCGCTATGTGCCCTAGCAGTACCGATCGTA 480
 |||||
 QY 1457 GCCCTATGAGAGGTTTCTGACAGTATAGTGGCAGCAGCCTCTTACACAAACCGAG 1516
 |||||
 Db 481 GCCCTATGAGAGGTTTCTGACAGTATAGTGGCAGCAGCCTCTTACACAAACCGAG 540
 |||||
 QY 1517 CAGTGCAGCAGCTCTGCGCACTGTAGAGGGCAGCAGTGGCCCTGAGAGCTGAG 1569
 |||||
 Db 541 CAGTGCAGCAGCTCTGCGCACTGTAGAGGGCAGCAGTGGCCCTGAGAGCTGAG 593
 |||||

RESULT 11
 BM795113 595 bp mRNA linear EST 05-MAR-2002
 LOCUS K-EST0076667 S21SNU520 Homo sapiens cDNA clone S21SNU520-27-E03 5',
 DEFINITION mRNA sequence.
 ACCESSION BM795113
 VERSION BM795113
 KEYWORDS
 SOURCE human,
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 595)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kr.ibm.re.kr
 Plate: 27 row: B column: 09
 High quality sequence stop: 595.
 Location/Qualifiers

FEATURES
 source
 1..595
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S21SNU520-27-E03"
 /clone_1ib="S21SNU520"
 /sex="F"


```
Db 143 GGGGCTATGTCCTCCCTTGGAGTACCGATTCGTAGCCCTTATGAGAAAGTTTCTGCAAGTA 84
|||
Oy 1484 ATGGTGGCAGCAGAGCCTCTCTTACACAAACCCAGACAGTGGACCACTTCTGCAACTTGG 1543
|||
Db 83 ATGGTGGCAGCAGCCTCTCTTACACAAACCCAGACAGTGGACCACTTCTGCAACTTGG 24
|||
Oy 1544 AGGGGCACTGCGCCCTCTGAGCT 1566
|||
Db 23 AGGGGCACTGCGCCCTCTGAGCT 1
|||

RESULT 13
BM793904 559 bp mRNA linear EST 05-MAR-2002
LOCUS K-BEST0074890 S22SNUI6n1 Homo sapiens cDNA clone S22SNUI6n1-32-A02
DEFINITION 5', mRNA sequence.
ACCESSION BM793904
VERSION BM793904.1 GI:19142136
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4470
Email: yongjung@mail.kribb.re.kr
Place: 32. row: A column: 02
High quality sequence stop: 559.
Location/Qualifiers
1..559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNUI6n1-32-A02"
/clone_1lb="S22SNUI6n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblastic-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; The S22SNUI6 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

BASE COUNT 120 a 154 c 145 g 140 t
ORIGIN
Query Match 35.4%; Score 555.8; DB 14; Length 559;
Best Local Similarity 99.6%; Pred. No. 3.9e-114;
Matches 557; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 944 CCTCTCGAAGATCCGACGACGACTACTACAGAGCTGACAGAGCACTTCTGAAA 1003
|||
Db 1 CCTCTCGAAGATCCGACGACGACTACTACAGAGCTGACAGAGCACTTCTGAAA 60
|||
Oy 1004 TGTTTTGCAGATTATTAACAAGGGGTTTTTCTGGGCTCTCCAAATATTAAGTTCAGGC 1063
|||
Db 61 TGTTTTGCAGATTATTAACAAGGGGTTTTTCTGGGCTCTCCAAATATTAAGTTCAGGC 120
|||
```

```
Oy 1064 CAGGATCTGTGTGTGTATCAATTGACTGTGGCTTCCGAGAAGTACATCATGTGCAGG 1123
|||
Db 121 CAGGATCTGTGTGTGTATCAATTGACTGTGGCTTCCGAGAAGTACATCATGTGCAGG 180
|||
Oy 1124 AGGTGAGACAGAGTTGATCAATGATATTAACAAGAGAGAGCTCTGTGATATTAAGTACGA 1183
|||
Db 181 AGGTGAGACAGAGTTGATCAATGATATTAACAAGAGAGAGCTCTGTGATATTAAGTACGA 240
|||
Oy 1184 TCTCAGAGCTCAGCGTGAAGTCAATGTGCAATTCCTTTCTCTGCCCAAGTCTGGGGTGGG 1243
|||
Db 241 TCTCAGAGCTCAGCGTGAAGTCAATGTGCAATTCCTTTCTCTGCCCAAGTCTGGGGTGGG 300
|||
Oy 1244 TGCAGAGCTGGGGGATGCGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1303
|||
Db 301 TGCAGAGCTGGGGGATGCGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 360
|||
Oy 1304 TCTATCTGATTTGCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1363
|||
Db 361 TCTATCTGATTTGCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 420
|||
Oy 1364 TCTTTTCAGCCCGGATACCTTACATCTCTATGACGAGTACCCCACTTACCAACCCATG 1423
|||
Db 421 TCTTTTCAGCCCGGATACCTTACATCTCTATGACGAGTACCCCACTTACCAACCCATG 480
|||
Oy 1424 GGGGCTATGTGCCCCCTTACGAGTACCGATGCTGTAACCCCTTATGAGAAGGTTTCTGCAAGTA 1483
|||
Db 481 GGGGCTATGTGCCCCCTTACGAGTACCGATGCTGTAACCCCTTATGAGAAGGTTTCTGCAAGTA 540
|||
Oy 1484 ATGTGTGAGCAGAGCTCTC 1502
|||
Db 541 ATGTGTGAGCAGAGCTCTC 559
|||

RESULT 14
LOCUS BG774910 981 bp mRNA linear EST 15-MAY-2001
DEFINITION 602649832P1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:4761054 5',
|||
ACCESSION BG774910
VERSION BG774910.1 GI:14045227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 981)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/BTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: I.M.G.E. Consortium distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Place: LDCM612 row: 0 column: 07
High quality sequence stop: 874.
Location/Qualifiers
1..981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4761054"
/clone_1lb="NIH MGC 40"
/tissue_type="Carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB1; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACAGAG(G). Library constructed by
```



```

Db 61 |||||
GACTGCGCTTCGAGAGGTACCATCATGTCCACAGCTGGAGACAGATTCAATCA 120

QY 1146 GTATAAACGGAAGAGCTCTCGATATATACTGACGATCTCAGACGTGAGTCA 1205
|||||

Db 121 GTATAAACGGAAGAGCTCTCGATATATACTGACGATCTCAGACGTGAGTCA 180
|||||

QY 1206 TGTGCAATTCCTTTCTCTGCGCATGTGCGGGCTGGGGTGCAGGCTGGGGCATGCGCT 1265
|||||

Db 181 TGTGCAATTCCTTTCTCTGCGCATGTGCGGGCTGGGGTGCAGGCTGGGGCATGCGCT 240
|||||

QY 1266 GCTGCTGCTGCTCTGTGTTCTGTGCGCTGGGCAATTGTCTATCTCATTTGCTTGGCTGT 1325
|||||

Db 241 GCTGCTGCTGCTGTGTGTTCTGTGCGCTGGGCAATTGTCTATCTCATTTGCTTGGCTGT 300
|||||

QY 1326 CTGTCAGTGCGCGGAAAGAACTACGCGGAGCTGGACATCTTTCAGCCGGGATACCTA 1385
|||||

Db 301 CTGTCAGTGCGCGGAAAGAACTACGCGGAGCTGGACATCTTTCAGCCGGGATACCTA 360
|||||

QY 1386 CCAATCCTATGAGCGAGTACCCCACTACACACCCCATGGGGGCTATGTGCCCTTAGCAG 1445
|||||

Db 361 CCAATCCTATGAGCGAGTACCCCACTACACACCCCATGGGGGCTATGTGCCCTTAGCAG 420
|||||

QY 1446 TACCGATGCTAGCCCTTATGAGAAGGTTCTGCAAGTAAATGTGGCAGCAGCTCTCTTA 1505
|||||

Db 421 TACCGATGCTAGCCCTTATGAGAAGGTTCTGCAAGTAAATGTGGCAGCAGCTCTCTTA 480
|||||

QY 1506 CACAAACCCAGAGTGGCAGCCACTTCTGCAACTTGTAGGGGACGTGCGCTCTGAGC 1565
|||||

Db 481 CACAAACCCAGAGTGGCAGCCACTTCTGCAACTTGTAGGGGACGTGCGCTCTGAGC 540
|||||

QY 1566 TTAGTGG 1572
|||||

Db 541 TTAGTGG 547
|||||

```

Search completed: May 8, 2003, 16:57:38
 Job time : 2462 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 15:26:48 ; Search time 766 Seconds
(without alignments)
9678.616 million cell updates/sec

Title: US-09-658-621B-1

Sequence: 1 gaattccctggctgcttgaa...tcgccctctgagctgagtgg 1572

Scoring table: IDENTITY_NUC

Searched: 6120518 seqs, 2358085216 residues

Total number of hits satisfying chosen parameters: 12241036

Maximum DB seq length: 20000000000

| | | |
|------------------|---------------|------|
| Post-processing: | Minimum Match | 0% |
| | Maximum Match | 100% |

Listing first 45 summaries

Database : Pending Patents NA New: *

```

1: 1/cnrg2_6/pdatodate/2/pna/US06_NEW_COMB.seq.*
2: 1/cnrg2_6/pdatodate/2/pna/US06_NEW_COMB.seq.*
3: 1/cnrg2_6/pdatodate/2/pna/US07_NEW_COMB.seq.*
4: 1/cnrg2_6/pdatodate/2/pna/US07_NEW_COMB.seq.*
5: 1/cnrg2_6/pdatodate/2/pna/US09_NEW_COMB.seq.*
6: 1/cnrg2_6/pdatodate/2/pna/US09_NEW_COMB.seq.*
7: 1/cnrg2_6/pdatodate/2/pna/US10_NEW_COMB.seq.*
8: 1/cnrg2_6/pdatodate/2/pna/US10_NEW_COMB.seq.*
9: 1/cnrg2_6/pdatodate/2/pna/US00_NEW_COMB.seq.*
10: 1/cnrg2_6/pdatodate/2/pna/US00_NEW_COMB.seq.*
11: 1/cnrg2_6/pdatodate/2/pna/US00_NEW_COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------------------|---------------------|
| 1 | 1572 | 100.0 | 1572 | 5 | US-09-658-621B-1 | Sequence 1, Appl1 |
| 2 | 1481.2 | 94.2 | 1721 | 9 | US-10-342-887-775 | Sequence 775, App |
| 3 | 1412.2 | 89.8 | 1804 | 9 | US-10-348-119-113 | Sequence 113, App |
| 4 | 1412.2 | 89.8 | 1804 | 10 | US-60-459-782-30 | Sequence 30, Appl |
| 5 | 1412.2 | 89.8 | 1804 | 10 | US-60-457-982-30 | Sequence 30, Appl |
| 6 | 1412.2 | 89.8 | 1823 | 9 | US-10-101-510-339 | Sequence 339, App |
| 7 | 1411.8 | 89.8 | 1549 | 8 | US-10-170-235-2491.5 | Sequence 2491.5, A |
| 8 | 1409.8 | 88.7 | 1549 | 11 | US-60-445-680-471 | Sequence 471, App |
| 9 | 1409.8 | 88.7 | 1549 | 11 | US-60-445-1335-343 | Sequence 443, App |
| 10 | 1409.8 | 88.7 | 1549 | 11 | US-60-445-050-343 | Sequence 343, App |
| 11 | 1409.2 | 75.7 | 4139 | 1 | PCIT-US92-19699A-310 | Sequence 310, App |
| 12 | 1353.8 | 60.6 | 1353 | 8 | US-10-170-225-24364 | Sequence 24364, App |
| 13 | 953.8 | 60.6 | 1353 | 11 | US-60-445-680-4774 | Sequence 4774, App |
| 14 | 952.2 | 60.6 | 1353 | 11 | US-60-445-1335-346 | Sequence 346, App |
| 15 | 952.2 | 60.6 | 1353 | 11 | US-60-445-050-345 | Sequence 345, App |
| 16 | 816.6 | 54.0 | 1272 | 8 | US-10-170-235-24365 | Sequence 24365, A |
| 17 | 817.8 | 53.9 | 1272 | 11 | US-60-445-680-473 | Sequence 473, App |
| 18 | 817.8 | 53.9 | 1272 | 11 | US-60-445-1335-345 | Sequence 345, App |
| 19 | 817.8 | 53.9 | 1272 | 11 | US-60-445-050-345 | Sequence 345, App |
| 20 | 750 | 47.7 | 8186 | 11 | US-60-442-116-21 | Sequence 21, Appl |
| 21 | 632.8 | 40.3 | 1173 | 8 | US-10-170-235-24366 | Sequence 24366, A |
| 22 | 631.2 | 40.2 | 1173 | 11 | US-60-445-680-472 | Sequence 472, App |

| | | | | | | |
|----|-------|------|------|----|---------------------|---------------------|
| 45 | 180.4 | 11.6 | 201 | 11 | US-60-452-680-29236 | Sequence 2233c, App |
| 44 | 159 | 12.7 | 201 | 11 | US-60-453-050-19350 | Sequence 1935d, App |
| 43 | 159 | 12.7 | 201 | 11 | US-60-453-050-19345 | Sequence 1934e, App |
| 42 | 159 | 12.7 | 201 | 11 | US-60-453-135-19350 | Sequence 1935a, App |
| 41 | 159 | 12.7 | 201 | 11 | US-60-453-135-19345 | Sequence 1934f, App |
| 40 | 159 | 12.7 | 201 | 11 | US-60-452-680-29232 | Sequence 2823c, App |
| 39 | 159 | 12.7 | 201 | 11 | US-60-452-680-29227 | Sequence 2823d, App |
| 38 | 200.6 | 12.8 | 201 | 11 | US-60-452-680-29237 | Sequence 2823e, App |
| 37 | 200.6 | 12.8 | 201 | 11 | US-60-453-135-19343 | Sequence 1934d, App |
| 36 | 200.6 | 12.8 | 201 | 11 | US-60-453-135-19333 | Sequence 1933d, App |
| 35 | 212.8 | 13.5 | 216 | 9 | US-10-283-017-941 | Sequence 941, App |
| 34 | 212.8 | 13.5 | 216 | 9 | US-10-283-017-941 | Sequence 941, App |
| 33 | 212.8 | 13.5 | 216 | 9 | US-10-283-017-941 | Sequence 941, App |
| 32 | 212.8 | 13.5 | 216 | 9 | US-10-283-017-941 | Sequence 941, App |
| 31 | 212.8 | 13.5 | 216 | 9 | US-10-283-017-941 | Sequence 941, App |
| 30 | 282.2 | 18.0 | 573 | 9 | US-10-101-510-049 | Sequence 9, App1 |
| 29 | 282.2 | 18.0 | 573 | 9 | US-10-101-510-049 | Sequence 9, App1 |
| 28 | 328.4 | 20.9 | 330 | 1 | PCT-US02-34777-1210 | Sequence 1210, App |
| 27 | 328.4 | 20.9 | 330 | 1 | PCT-US02-34777-1210 | Sequence 1210, App |
| 26 | 328.4 | 20.9 | 330 | 1 | PCT-US02-34777-1210 | Sequence 1210, App |
| 25 | 328.4 | 20.9 | 330 | 1 | PCT-US02-34777-1210 | Sequence 1210, App |
| 24 | 328.4 | 20.9 | 330 | 1 | PCT-US02-34777-1210 | Sequence 1210, App |
| 23 | 631.2 | 40.2 | 1173 | 11 | US-60-453-135-1944 | Sequence 344, App |
| 22 | 631.2 | 40.2 | 1173 | 11 | US-60-453-135-1944 | Sequence 344, App |
| 21 | 544 | 37.2 | 216 | 9 | US-10-101-510-336 | Sequence 326, App |
| 20 | 544 | 37.2 | 216 | 9 | US-10-101-510-336 | Sequence 326, App |
| 19 | 544 | 37.2 | 216 | 9 | US-10-101-510-336 | Sequence 326, App |
| 18 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 17 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 16 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 15 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 14 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 13 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 12 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 11 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 10 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 9 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 8 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 7 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 6 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 5 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 4 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 3 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 2 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |
| 1 | 459 | 29.2 | 5660 | 9 | US-10-311-455-2330 | Sequence 2330, App |

ALIGNMENTS

```

1 RESULT 1
2 US-09-658-621B-1
3 Sequence 1, Application US/09658621B
4 GENERAL INFORMATION:
5 APPLICANT: Taylor-Papadimitriou, Joyce
6 APPLICANT: Heukamp, Lukas Carl
7 APPLICANT: Officina, Rieck
8 APPLICANT: Melief, Cornelis Johanna Maria
9 APPLICANT: Acres, Bruce
10 APPLICANT: Thomas, Mireille
11 TITLE OF INVENTION: WOC-1 derived peptides
12 FILE REFERENCE: 029395-017
13 CURRENT APPLICATION NUMBER: US/09/658,621B
14 CURRENT FILING DATE: 2000-09-08
15 PRIOR APPLICATION NUMBER: US 60/187,215
16 PRIOR FILING DATE: 2000-03-03
17 PRIOR APPLICATION NUMBER: GB 9921242.5
18 PRIOR FILING DATE: 1999-09-08
19 PRIOR APPLICATION NUMBER: EP 99 40 2237.4
20 PRIOR FILING DATE: 1999-09-10
21 NUMBER OF SEQ ID NOS: 80
22 SOFTWARE: PatentIn version 3.1

```

```

? SEQ ID NO 1
? LENGTH: 1572
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (58) ..(1542)
?
US-09-658-621B-1

```

| | | | | |
|------------------------|---------|-------------|---------|--------------|
| Query Match | 100.0%; | Score 1572; | DB 5; | Length 1572; |
| Post-Test 1 Similarity | 100.0%; | Dread No. 7 | 1e-396; | |

Matches 1572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Oy | 1 | GAATTCCTGCTGCTGTAATCTGTTCTGCCCCCTTCCACCATTTCACCAACCAACAG | 60 |
| | | | |
| Db | 1 | GAATTCCTGCTGCTGTAATCTGTTCTGCCCCCTTCCACCATTTCACCAACCAACAG | 60 |
| Oy | 61 | AACACGGGCAACCCAGTCTCTTTCTTCTGCTGCTGCTCTCAACAGTGTCAAGTGT | 120 |
| Db | 61 | AACACGGGCAACCCAGTCTCTTTCTTCTGCTGCTGCTCTCAACAGTGTCAAGTGT | 120 |
| Oy | 121 | ACAGGTTTGTGTATGCAAGCTTACCCCAAGTGGAGAAAGAGATTTCGGCTACCAAG | 180 |
| | | | |

```
Db 121 ACAGGTTCTGTCATGCAAGCTTACCCAGGTGAGAAAAGAGACCTTGAGTACCAG 180
Oy 181 AGAAGTTCCAGTCCAGGCTTACTGAGAGAAATGCTGTGAGATATGACAGAGGCTACTC 240
Db 181 AGAAGTTCCAGTCCAGGCTTACTGAGAGAAATGCTGTGAGATATGACAGAGGCTACTC 240
Oy 241 TCCAGCCACAGCCCCGGGTTCAAGGCTCTTCCACACTCAGGGACAGAGTGTCACTTGACC 300
Db 241 TCCAGCCACAGCCCCGGGTTCAAGGCTCTTCCACACTCAGGGACAGAGTGTCACTTGACC 300
Oy 301 CCGGACACAGGAAACAGCTTTCAGTTCAGCTGCACCTGAGGACAGAGTGTCACTTGACC 360
Db 301 CCGGACACAGGAAACAGCTTTCAGTTCAGCTGCACCTGAGGACAGAGTGTCACTTGACC 360
Oy 361 CCAAGTACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 420
Db 361 CCAAGTACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 420
Oy 421 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 480
Db 421 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 480
Oy 481 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 540
Db 481 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 540
Oy 541 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 600
Db 541 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 600
Oy 601 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 660
Db 601 CCGGACACAGGACAGGCTTGGGCTCCACACCCGACAGCCACAGTGTCACTTGACC 660
Oy 661 TCAAGGCTCTGCAATCAGGCTCAGCTTCTACTCTGTGTGCAAAACAGGACCTTGACC 720
Db 661 TCAAGGCTCTGCAATCAGGCTCAGCTTCTACTCTGTGTGCAAAACAGGACCTTGACC 720
Oy 721 ACCACAAACCCAGCCAGCAGACAGCTCCACCCAGCATTTCCACAGCAGCAGCTTGACT 780
Db 721 ACCACAAACCCAGCCAGCAGACAGCTCCACCCAGCATTTCCACAGCAGCAGCTTGACT 780
Oy 781 CCAACACACCTTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 840
Db 781 CCAACACACCTTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 840
Oy 841 GTAACCTCTCTCACTCTCTCAATCAGAGACTTCTCCCAAGTTCTCTCTGAGAGATCCC 900
Db 841 GTAACCTCTCTCACTCTCTCAATCAGAGACTTCTCCCAAGTTCTCTCTGAGAGATCCC 900
Oy 901 TTCTTTTCTCTGTCTTTTCAATTTCAATTTCAATTTCAATTTTCTCTCTGAGAGATCCC 960
Db 901 TTCTTTTCTCTGTCTTTTCAATTTCAATTTCAATTTTCTCTCTGAGAGATCCC 960
Oy 961 AGCAGCGACCTACTACCAAGAGCTGACAGAGAGACATTTTGGAAATGTTTGGACATTTAT 1020
Db 961 AGCAGCGACCTACTACCAAGAGCTGACAGAGAGACATTTTGGAAATGTTTGGACATTTAT 1020
Oy 1021 AAACAGAGGGGTTTCTGTGGGCTCTCCAAATTAATTAAGTTCAAGCCAGATTTGGTGGTA 1080
Db 1021 AAACAGAGGGGTTTCTGTGGGCTCTCCAAATTAATTAAGTTCAAGATTTGGTGGTA 1080
Oy 1081 CAATTGACTCTGAGCTTCCAGAGAGTACCAATGATGTCACAGAGTGTGACACAGTTTC 1140
Db 1081 CAATTGACTCTGAGCTTCCAGAGAGTACCAATGATGTCACAGAGTGTGACACAGTTTC 1140
Oy 1141 AATCAGATTAATAAGAGAGAGAGCTCTCCAGATTAATACTGAGATCTCAGAGCTGAGCGTG 1200
Db 1141 AATCAGATTAATAAGAGAGAGAGCTCTCCAGATTAATACTGAGATCTCAGAGCTGAGCGTG 1200
Oy 1201 AGTCAATGTCATTTCTTTTCTGTGACCACTGTGGGCTGTGGGCTGTGGGCTGTGGGCTATC 1260
Db 1201 AGTCAATGTCATTTCTTTTCTGTGACCACTGTGGGCTGTGGGCTGTGGGCTGTGGGCTATC 1260
```

```
Oy 1261 GCGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1320
Db 1261 GCGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1320
Oy 1321 GCTGTCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1380
Db 1321 GCTGTCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1380
Oy 1381 ACTTACCATCTTATGAGGAGTACCCACCTTACACACACACACACACACACACACACACACAC 1440
Db 1381 ACTTACCATCTTATGAGGAGTACCCACCTTACACACACACACACACACACACACACACACAC 1440
Oy 1441 AGCAGTACCGATCTGTAGCCCTTATGAGAGATTTGTGAGGTAATGTTGACAGACACTTC 1500
Db 1441 AGCAGTACCGATCTGTAGCCCTTATGAGAGATTTGTGAGGTAATGTTGACAGACACTTC 1500
Oy 1501 TCTTACACAAACCCAGCAGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
Db 1501 TCTTACACAAACCCAGCAGTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
Oy 1561 TGAAGTGAATGG 1572
Db 1561 TGAAGTGAATGG 1572
```

```
RESULT 2
US-10-342-887-775
; Sequence 775, Application US/10342887
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 775
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-775
```

```
Query Match 94.2%; Score 1481.2; DB 9; Length 1721;
Best Local Similarity 95.5%; Pred. No. 31e-278;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;
```

```
Oy 1  GAATTCCTGCTGCTGTTGAATCTGTTCTGCCCCCTCCACCCCATTTTCAACACACACATG 60
Db 1  GAATTCCTGCTGCTGTTGAATCTGTTCTGCCCCCTCCACCCCATTTTCAACACACACATG 60
Oy 61  ACACCGGGACCCAGCTCTCTTTCTTCTGCTGCTGCTCTCCACAGTGTCAAGTTGTT 120
Db 61  ACACCGGGACCCAGCTCTCTTTCTTCTGCTGCTGCTCTCCACAGTGTCAAGTTGTT 120
Oy 121  ACAAGTTCTGTGATGCAAGCTTACCCAGGAGGAGAAAGAGACCTTGGGCTACCCAG 180
Db 121  ACAAGTTCTGTGATGCAAGCTTACCCAGGAGGAGAAAGAGACCTTGGGCTACCCAG 180
Oy 181  AGAAGTTCAATGCTTCACTGAGAGAGATGCTGAGTATGACCAAGCAGCTACTC 240
Db 181  AGAAGTTCAATGCTTCACTGAGAGAGATGCTGAGTATGACCAAGCAGCTACTC 240
```


[illegible]

```

OY 1450 GATCGACACCCCTTATGAAAGTTTCTCAGATTAATGTGGACAGACGCTCTTTACACA 1509
Db 1405 GATCTGATGCCCTTATGAAAGTTTCTGAGATTAATGTGGACAGACGCTCTTTACACA 1464
OY 1510 AACCCAGCAGTGGCAGCACCATTGTGCACATTGTAGGGGCGACGTCGCTCTTGAAGTGA 1569
Db 1465 AACCCAGCAGTGGCAGCACCATTGTGCACATTGTAGGGGCGACGTCGCTCTTGAAGTGA 1524
OY 1570 TGG 1572
Db 1525 TGG 1527

RESULT 4
/ Sequence 30, Application US/60459782
/ GENERAL INFORMATION:
/ APPLICANT: Wyeth Research
/ APPLICANT: Twine, Natalie C.
/ APPLICANT: Stover, Jennifer A.
/ APPLICANT: Trepicchio, William L.
/ APPLICANT: Dornez, Andrew J.
/ APPLICANT: Slonim, Donna K.
/ APPLICANT: Buczynski, Michael E.
/ TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND/OR SOLID TUMORS
/ FILE REFERENCE: AM101055
/ CURRENT APPLICATION NUMBER: US/60459,782
/ NUMBER OF SEQ ID NOS: 331
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 30
/ LENGTH: 1804
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-60-459-782-30

Query Match 89.8%; Score 1412.2; DB 10; Length 1804;
Best Local Similarity 95.3%; Pred. No. 7.9e-265;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps

OY 10 GGCTGTTGATCTGTTCTGCGCCGCTCCGCCACCATTTCCACACACCATGACACGGGGC 69
Db 25 GGCTGCTGAATCTGTTCTGCGCCGCTCCGCCACCATTTCCACACCATGACACGGGGC 84
OY 70 ACCCAGTCTCTTCTTCTCTGCTGCTCTCTCAAGTGTCTTACAGTGTCT 129
Db 85 ACCCAGTCTCTTCTTCTCTGCTGCTCTCTCAAGTGTCTTACAGTGTCT 144
OY 130 GGTATGCAACTTACCCAGGATGAGAAAGAGACTTCGGCTACCAAGAGAATTCA 189
Db 145 GGTATGCAACTTACCCAGGATGAGAAAGAGACTTCGGCTACCAAGAGAATTCA 204
OY 190 GTGCCCAAGTCTACGAGAAAGATGCTGATGATGACGACGAGGTACTCTCCAGGAC 249
Db 205 GTGCCCAAGTCTACGAGAAAGATGCTGATGATGACGAGGTACTCTCCAGGAC 264
OY 250 AGCTCCGTTTACGCTCTTCAACAACCTAGAGAAAGAGATGTCACTGTGACCCCGACAG 309
Db 265 AGCTCCGTTTACGCTCTTCAACAACCTAGAGAAAGAGATGTCACTGTGACCCCGACAG 324
OY 310 GAACCAAGTTCAGTTCAGTTCGACCTTGGGAGACAGGATGTCACTCGTCTCCAGTACC 369
Db 325 GAACCAAGTTCAGTTCAGTTCGACCTTGGGAGACAGGATGTCACTCGTCTCCAGTACC 384
OY 370 AGGCCAGGCTGGGCTTCAACACCCCGACCCAGATGATACCTCAAGCCCGGAGAC 429
Db 385 AGGCCAGGCTGGGCTTCAACACCCCGACCCAGATGATACCTCAAGCCCGGAGAC 444
OY 430 AAGCTAGCCCTGGGCTTCAACGCTCCCGGACCAAGTGTCACTCGGCTCCGAGACAC 489
Db 445 AAGCTAGCCCTGGGCTTCAACGCTCCCGGACCAAGTGTCACTCGGCTCCGAGACAC 474

```

Thu May 8 18:53:10 2003

us-09-658-621b-1.11

Page 5

[illegible]

```

Db      1525 TGG 1527
|||||
RESULT 5
US-60-427-982-30
; Sequence 10 Application US/60427982
; GENERAL INFORMATION:
; APPLICANT: Wyeth Research
; APPLICANT: Butczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dornier, Andrew
; APPLICANT: Treplechio, William
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND/OR SOLID TUMORS
; FILE REFERENCE: AM101055
; CURRENT APPLICATION NUMBER: US/60/427,982
; CURRENT FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-427-982-30

Query Match      89.8%; Score 1412.2; DB 11; Length 1804;
Best Local Similarity 95.3%; Pred No 7.9e-265;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps

1:
10 GCGCTGTTGAATCTGTTCTGTCGCCCTCTCCCAACCAATTTCACCAACCAACATGACACCGGCG 69
Db      25 GCGCTGCTTAATCTGTTCTGTCGCCCTCTCCCAACCAATTTCACCAACCAACATGACACCGGCG 84
QY      70 ACCCACTCTCTTCTTCTCTGCTGCTGCTCTCTCAAGTGCTTCAAGTTGTTCAGGTTCT 129
Db      85 ACCCACTCTCTTCTTCTCTGCTGCTGCTCTCTCAAGTGCTTCAAGTTGTTCAGGTTCT 144
QY      130 GGTTCATGCAAGCTCTACCCCAAGGTGAGAAAGAGAAGAGACTTCGCGCTACCCAGAGAATTCA 189
Db      145 GGTTCATGCAAGCTCTACCCCAAGGTGAGAAAGAGAAGAGACTTCGCGCTACCCAGAGAATTCA 204
QY      190 GTGCCAGCTCTACTGAGAAAGATGCTGTGAGATGTGACACGACGACCTATCTCCAGGCAC 249
Db      205 GTGCCAGCTCTACTGAGAAAGATGCTGTGAGATGTGACACGACGACCTATCTCCAGGCAC 264
QY      250 AGCCCGGTTTGAAGCTCTCTCAACAATCAAGAGACAGATGTCACTCTGGCCCGGCGACG 309
Db      265 AGCCCGGTTTGAAGCTCTCTCAACAATCAAGAGACAGATGTCACTCTGGCCCGGCGACG 324
QY      310 GAACCAAGCTTCAAGTTTCAAGTGTCCACCTTGGGGAACAGATGTCACTCTGGTCCAGTCAAC 369
Db      325 GAACCAAGCTTCAAGTTTCAAGTGTCCACCTTGGGGAACAGATGTCACTCTGGTCCAGTCAAC 384
QY      370 AAGCGAAGCTCTGAGGCTTCCACCAACCCCGGACGACCCAGATGTCACTCAAGCCCGGACAC 429
Db      385 AAGCGAAGCTCTGAGGCTTCCACCAACCCCGGACGACCCAGATGTCACTCAAGCCCGGACAC 444
QY      430 AAGCGAAGCTCTGAGGCTTCCACCGGCCCCCGGCGCCACAGATGTCACTCTGGCCCGGACAC 489
Db      445 AAGCGAAGCTCTGAGGCTTCCACCGGCCCCCGGCGCCACAGATGTCACTCTGGCCCGGACAC 474
QY      490 AAGCGGCGCCCGGAGCTCCACCGGCCCCCGGCGCCACAGATGTCACTCTGGCCCGGACAC 539
Db      475 -----GCTCAAGGTGTCACTCTGGCCCGGACAC 504
QY      550 AAGCGGCGCCCGGAGCTCCACCGGCCCCCGGCGCCACAGATGTCACTCTGGCCCGGACAC 609
Db      505 AAGCGGCGCCCGGAGCTCCACCGGCCCCCGGCGCCACAGATGTCACTCTGGCCCGGACAC 564
QY      610 AAGCGGCGCTTGGGCTTCAACCGGCTCTCAAGTTTCAAGATGTCACTCTGGCCCGGACGCTCT 669
Db      565 AAGCGGCGCTTGGGCTTCAACCGGCTCTCAAGTTTCAAGATGTCACTCTGGCCCGGACGCTCT 624

```

```
Qy 670 GCATCAGGCTCAGGCTTCTACTCTGNGCAAGAGGCACTTGGCAGGGCTACCAAGC 729
Db 625 GATTCAGGCTCAGGCTTCTACTCTGNGCAAGAGGCACTTGGCAGGGCTACCAAGC 684
Qy 730 CCAGCCAGAGAGAGCATCTCCAGCCAGATTCGACCAAGCACTTGATCTCTACCAAC 789
Db 685 CCAGCCAGAGAGAGCATCTCCAGCCAGATTCGACCAAGCACTTGATCTCTACCAAC 744
Qy 790 CTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 849
Db 745 CTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 804
Qy 850 CTCACCTCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 909
Db 805 CTCACCTCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 864
Qy 910 CTGCTCTTTCAGATTTGAAACCTCCAGTTTATTCCTCTGAGAGATCCAGACCAAC 969
Db 865 CTGCTCTTTCAGATTTGAAACCTCCAGTTTATTCCTCTGAGAGATCCAGACCAAC 924
Qy 970 TACTACAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
Db 925 TACTACAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 984
Qy 1030 GGTTCCTGAGGCTCTCTCAATATTAATTAAGTTCAGGCTAGATCTGAGTGTGCA 1089
Db 985 GGTTCCTGAGGCTCTCTCAATATTAATTAAGTTCAGGCTAGATCTGAGTGTGCA 1044
Qy 1090 CTGGCTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
Db 1045 CTGGCTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104
Qy 1150 AAAACGAGAGAGAGCTCTGATTAACCTGACGATCTCAGAGCTCAGAGTATCATG 1209
Db 1105 AAAACGAGAGAGAGCTCTGATTAACCTGACGATCTCAGAGCTCAGAGTATCATG 1164
Qy 1210 CCATTTCTCTCTCTGAGCAGCTGAGGAGTGGAGTGGAGAGAGAGAGAGAGAG 1269
Db 1165 CCATTTCTCTCTCTGAGCAGCTGAGGAGTGGAGTGGAGAGAGAGAGAGAGAG 1224
Qy 1270 GTGCTGCTGTGTGTTGTGTTGCTGAGCAGTGTGCTGATCTCATTTGCTGTGT 1329
Db 1225 GTGCTGCTGTGTGTTGTGTTGCTGAGCAGTGTGCTGATCTCATTTGCTGTGT 1284
Qy 1330 CAGTGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1389
Db 1285 CAGTGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
Qy 1390 CCTATGAGAGAGTACCCCACTTACAGACCCATGAGGCTATGTGCTCTAGAGTAC 1449
Db 1345 CCTATGAGAGAGTACCCCACTTACAGACCCATGAGGCTATGTGCTCTAGAGTAC 1404
Qy 1450 GATCGAGAGCCTATGAGAGAGTCTGAGAGTATAGTGGAGAGAGCTCTCTAGAC 1509
Db 1405 GATCGAGAGCCTATGAGAGAGTCTGAGAGTATAGTGGAGAGAGCTCTCTAGAC 1464
Qy 1510 AACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1569
Db 1465 AACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
Qy 1570 TGG 1572
Db 1525 TGG 1527
```

```
RESULT 6
US-10-101-510-339
: Sequence 339, Application US/10101510
: GENERAL INFORMATION:
: APPLICANT: WAN, JACKSON
: APPLICANT: WANG, YIXIN
: TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
: FILE REFERENCE: 15117.0012

CURRENT APPLICATION NUMBER: US/10/101,510
: CURRENT FILING DATE: 2002-03-20
: PRIOR APPLICATION NUMBER: 60/276,947
: PRIOR FILING DATE: 2001-03-20
: NUMBER OF SEQ ID NOS: 805
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 339
: LENGTH: 1823
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-101-510-339

Query Match      89.8%; Score 1412.2; DB 9; Length 1823;
Best Local Similarity 95.3%; Pred. No. 7.9e-265;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps 1;
```

```
Qy 10 GGTGCTTGAACTGTTCTGTCGCCCCCTCCACCATTTGACAGCAGCAGCAGGAG 69
Db 25 GCTGCTGAACTGTTCTGTCGCCCCCTCCACCATTTGACAGCAGCAGCAGGAG 84
Qy 70 ACCGATCTCTTTCTTCTGCTGCTGCTCTCAAGTCTTACAGTTGTTACAGTTCT 129
Db 85 ACCGATCTCTTTCTTCTGCTGCTGCTCTCAAGTCTTACAGTTGTTACAGTTCT 144
Qy 130 GGTATGCAAGCTCTACCCAGAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAG 189
Db 145 GGTATGCAAGCTCTACCCAGAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAG 204
Qy 190 GTGGCCAGCTCTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249
Db 205 GTGGCCAGCTCTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 264
Qy 250 AGCCCGGATGAGGCTCTCTCCAGCAGTCAAGAGAGAGAGAGAGAGAGAGAG 309
Db 265 AGCCCGGATGAGGCTCTCTCCAGCAGTCAAGAGAGAGAGAGAGAGAGAGAG 324
Qy 310 GAACAGGTTAGAGTTACAGTCTGAGCAGCTGGAGAGAGAGAGAGAGAGAG 369
Db 325 GAACAGGTTAGAGTTAGAGTCTGAGCAGCTGGAGAGAGAGAGAGAGAGAGAG 384
Qy 370 AGGCGAGCTCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 429
Db 385 AGGCGAGCTCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 444
Qy 430 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 489
Db 445 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 474
Qy 490 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 549
Db 475 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 504
Qy 550 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 609
Db 505 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 564
Qy 610 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 669
Db 565 AAGCGAGCCCGAGGCTTCCAGCAGCCGAGAGAGAGAGAGAGAGAGAGAGAG 624
Qy 670 GCATCAGGCTCAGGCTTCTACTCTGAGTGAAGAGAGAGAGAGAGAGAGAG 729
Db 625 GCATCAGGCTCAGGCTTCTACTCTGAGTGAAGAGAGAGAGAGAGAGAGAG 684
Qy 730 CCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
Db 685 CCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 744
Qy 790 CTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 849
Db 745 CTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 804
Qy 850 CTCACCTCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 909
```



```

|||||
Db 805 CTACCTCTCCAAACAGACACTTCCCGAGTGTCTACCTGGAGCTCTTCTTTTC 864
Qy 910 CTGCTTTTACATTTCAAACTCCAGTTTAAATTCCTCTCTGGAAGATCCAGACCGAC 969
Db 865 CTGCTTTTCAATTTCAAACTCCAGTTTAAATTCCTCTCTGGAAGATCCAGACCGAC 924
Qy 970 TACTACCAAGAGCTGAGAGATTTCTGAAATGTTTGGAGATTTTAAACAAGG 1029
Db 925 TACTACCAAGAGCTGAGAGATTTCTGAAATGTTTGGAGATTTTAAACAAGG 984
Qy 1030 GGTTCCTGAGCTCTCCAAATATTTAAGTTCAAGCCAGATCTGTGTGATCAATTGACT 1089
Db 985 GGTTCCTGAGCTCTCCAAATATTTAAGTTCAAGCCAGATCTGTGTGATCAATTGACT 1044
Qy 1090 CTGAGCTTCCAGAGATGATCATATGTCAGAGCTGAGAGACAGTTCAATCATAT 1149
Db 1045 CTGAGCTTCCAGAGATGATCATATGTCAGAGCTGAGAGACAGTTCAATCATAT 1104
Qy 1150 AAAACGGAAGAGCTCTCCATATTTAATCTGAGATCTCAAGCTCAAGCTCATATG 1209
Db 1105 AAAACGGAAGAGCTCTCCATATTTAATCTGAGATCTCAAGCTCAAGCTCATATG 1164
Qy 1210 CCAATTCCTTTCTCTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1269
Db 1165 CCAATTCCTTTCTCTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1224
Qy 1270 GTGCTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1329
Db 1225 GTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1284
Qy 1330 CAGTCCGCGGAAAGACTAGAGGAGCTGAGATTTTCCAGCCCGGAAATCAATCAAT 1389
Db 1285 CAGTCCGCGGAAAGACTAGAGGAGCTGAGATTTTCCAGCCCGGAAATCAATCAAT 1344
Qy 1390 CCTATGAGAGATACCCCACTTACAGACCCATGAGGCTATATGTCCTTACAGATAC 1449
Db 1345 CCTATGAGAGATACCCCACTTACAGACCCATGAGGCTATATGTCCTTACAGATAC 1404
Qy 1450 GATCGTAGGCCCCATGAGAGATTTTCTGAGGATTAAGTGAGAGAGAGCTCTTACACA 1509
Db 1405 GATCGTAGGCCCCATGAGAGATTTTCTGAGGATTAAGTGAGAGAGAGCTCTTACACA 1464
Qy 1510 AACCCAGAGTGGAGCACTTCTGCAACTTGTAGAGGAGCACTGAGCTTCTGAGCTGAG 1569
Db 1465 AACCCAGAGTGGAGCACTTCTGCAACTTGTAGAGGAGCACTGAGCTTCTGAGCTGAG 1524
Qy 1570 TGG 1572
Db 1525 TGG 1527

```

```

RESULT 7
US-10-170-235-24915
; Sequence 24915, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; TITLE OF INVENTION: TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
; RITE REFERENCE: C1001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; NUMBER OF SEQ ID NOS: 2003-03-17
; SEQ ID NO 24915
; LENGTH: 1549
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-24915

```

```

Query Match 89.8%; Score 1411.8; DB 8; Length 1549;
Best Local Similarity 95.4%; Pred. No. 9,2e-265;
Matches 1485; Conservative 0; Mismatches 12; Indels 60; Gaps 1;

```

```

Qy 12 CTGCTTGAATGTGTCTGCCCCCTCCCAACCATTTTACACACCACTAGACACCGGAC 71
Db 1 CTGCTTGAATGTGTCTGCCCCCTCCCAACCATTTTACACACCACTAGACACCGGAC 60
Qy 72 CCAGTCTCTTCTCTCTGCTGCTGCTCTCTCAAGTGTCTTACAGTGTCTG 131
Db 61 CCAGTCTCTTCTCTCTGCTGCTGCTCTCTCAAGTGTCTTACAGTGTCTG 120
Qy 132 TCATGAGAGCTTACCCAGAGTGAAGAAAGAGACTTCCGCTACCCAGAGAGTTCACT 191
Db 121 TCATGAGAGCTTACCCAGAGTGAAGAAAGAGACTTCCGCTACCCAGAGAGTTCACT 180
Qy 192 GCCAGCTTCTAGAGAGATCTGTGATGATGACAGAGGATCTCTCAAGCAAG 251
Db 181 GCCAGCTTCTAGAGAGATCTGTGATGATGACAGAGGATCTCTCAAGCAAG 240
Qy 252 CCCCAGTTCAGGCTCTTCCACCACTAGAGGAGAGAGATGTCATCTGAGCCCGACGGA 311
Db 241 CCCCAGTTCAGGCTCTTCCACCACTAGAGGAGAGAGATGTCATCTGAGCCCGACGGA 300
Qy 312 ACCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGCTCCAGTCAAG 371
Db 301 ACCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGCTCCAGTCAAG 360
Qy 372 GCCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGAGCCCGACGGA 431
Db 361 GCCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGAGCCCGACGGA 420
Qy 432 GCCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGAGCCCGACGGA 491
Db 421 GCCAGCTTGAAGTTCACTGCTCACTGAGGAGAGAGATGTCATCTGAGCCCGACGGA 448
Qy 492 GCCGCCCCGAGCTTCAACCGCCCGCCGCAAGTGTCACTGAGCCCGGACACAG 551
Db 449 -----GCCGCCCCGAGCTTCAACCGCCCGCCGCAAGTGTCACTGAGCCCGGACACAG 480
Qy 552 GCCGCCCCGAGCTTCAACCGCCCGCCGCCGCAAGTGTCACTGAGCCCGGACACAG 611
Db 481 GCCGCCCCGAGCTTCAACCGCCCGCCGCCGCAAGTGTCACTGAGCCCGGACACAG 540
Qy 612 GCCGCCCCGAGCTTCAACCGCCCGCCGCCGCAAGTGTCACTGAGCCCGGACACAG 671
Db 541 GCCGCCCCGAGCTTCAACCGCCCGCCGCCGCCGCAAGTGTCACTGAGCCCGGACACAG 600
Qy 672 ATCAGAGCTCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAACCCC 731
Db 601 ATCAGAGCTCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAACCCC 660
Qy 732 AGCCAGAGAGAGCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAAC 791
Db 661 AGCCAGAGAGAGCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAAC 720
Qy 792 TGCAGAGAGAGCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAAC 851
Db 721 TGCAGAGAGAGCACTTCACTGTGTGACAAACGAGCACTTGTGAGAGGCTTACCAAC 780
Qy 852 CACCTCTCAATCAACAGCACTTCTCCAGTTGTCTACTGAGGCTCTCTTCTTCTTCT 911
Db 781 CACCTCTCAATCAACAGCACTTCTCCAGTTGTCTACTGAGGCTCTCTTCTTCTTCT 840
Qy 912 GTCTTTTCAATTTAAACCTCCAGTTTAAATCTCTGAGAGATCCAGAGAGCACTA 971
Db 841 GTCTTTTCAATTTAAACCTCCAGTTTAAATCTCTGAGAGATCCAGAGAGCACTA 900
Qy 972 CTACCAAGAGCTGAGAGAGCACTTCTGAGAAATGTTTTTGGAGATTTTAAACAAGGAG 1031
Db 901 CTACCAAGAGCTGAGAGAGCACTTCTGAGAAATGTTTTTGGAGATTTTAAACAAGGAG 960
Qy 1032 TTTTCTGAGGCTCTCCATATTTAAGTTCAAGGAGAGATCTGTGTGTGATGATCT 1091
Db 961 TTTTCTGAGGCTCTCCATATTTAAGTTCAAGGAGAGATCTGTGTGTGATGATCT 1020
Qy 1092 GGCCTTCCAGAGAGTCAATCAATGTCCACAGAGCTGAGAGCAAGTTCAATCAATATA 1151

```

[illegible]

| | | | | |
|---|---|------|---|------|
| D | b | 181 | GCACGACTTACAGAAAGAAATGCTGTGAGTATGACACGAGAGGTACTCTCCACGACAG | 240 |
| O | y | 252 | CCCCCGTTGAGGCTCTCCACCACTTCAGGGACAGAAATGTCACTCTGTGGCCCGGACAGAA | 311 |
| D | b | 241 | CCCCGGTTGAGGCTCTCCACCACTTCAGGGACAGAAATGTCACTCTGTGGCCCGGACAGAA | 300 |
| O | y | 312 | ACCACTTCAGATTTCAGCTGCACCTTGGGACAGAAATGTCACTCTGGTCCAGTCAACAG | 371 |
| D | b | 301 | ACCACTTCAGATTTCAGCTGCACCTTGGGACAGAAATGTCACTCTGGTCCAGTCAACAG | 360 |
| O | y | 372 | GCACGCTCTGGGCTCCACCACTCCGACAGCCAGATGTCACTCAAGCCCGGACAAACA | 431 |
| D | b | 361 | GCACGCTCTGGGCTCCACCACTCCGACAGCCAGATGTCACTCAAGCCCGGACAAACA | 420 |
| O | y | 432 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 491 |
| D | b | 421 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 448 |
| O | y | 492 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 551 |
| D | b | 449 | -----GCCACAGATGTCACTCGGCCCCGACACAG | 480 |
| O | y | 552 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 611 |
| D | b | 481 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 540 |
| O | y | 612 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 671 |
| D | b | 541 | GCACGCTCTGGGCTCCACCTGGCTCCCGGACAGATGTCACTCGGCCCCGACACAG | 600 |
| O | y | 672 | ATCAGGCTCAGCTTCTACTGTGTGACCAACGAGCACTGTGCAGAGGGCTACCAACCC | 731 |
| D | b | 601 | ATCAGGCTCAGCTTCTACTGTGTGACCAACGAGCACTGTGCAGAGGGCTACCAACCC | 660 |
| O | y | 732 | AGCAGCAAGAGACCTCCACCCAGCAATTCGACGCAACAATCTGATCTTACCAACCT | 791 |
| D | b | 661 | AGCAGCAAGAGACCTCCACCTGATTCATTCGACGCAACAATCTGATCTTACCAACCT | 720 |
| O | y | 792 | TGCACGCTTACGACCAACAGATGTGCAGTATGACCTCACATATGACGGTACTCTCT | 851 |
| D | b | 721 | TGCACGCTTACGACCAACAGATGTGCAGTATGACCTCACATATGACGGTACTCTCT | 780 |
| O | y | 852 | CACGCTCTCCAAATCAGACACTTCTCCGACGATGTGTACTAGGGGCTCTTTCTTTTCC | 911 |
| D | b | 781 | CACGCTCTCCAAATCAGACACTTCTCCGACGATGTGTACTAGGGGCTCTTTCTTTTCC | 840 |
| O | y | 912 | GTCTTTTCAATTTGAAACGTCCAGATTTATTCCTCTGTGGAAGATCCCAACACCACTA | 971 |
| D | b | 841 | GTCTTTTCAATTTGAAACGTCTCCAGATTTATTCCTCTGTGGAAGATCCCAACACCACTA | 900 |
| O | y | 972 | CTACCAAGCTGACAGAGACATTTCTGTAATGTTTTSCAGATTTTAAACAAGGGG | 1031 |
| D | b | 901 | CTACCAAGCTGACAGAGACATTTCTGTAATGTTTTSCAGATTTTAAACAAGGGG | 960 |
| O | y | 1032 | TTTTCTGGGCTCTCCAAATTTAAGTTACGAGGCAAGATCTGTGTGTGACATTTGACTCT | 1091 |
| D | b | 961 | TTTTCTGGGCTCTCCAAATTTAAGTTACGAGGCAAGATCTGTGTGTGACATTTGACTCT | 1020 |
| O | y | 1092 | GGCTCTTCGAGAGGTAACATCAATGTCCACGACGTGTGAACAAGTTCAATCAATTA | 1155 |
| D | b | 1021 | GGCTCTTCGAGAGGTAACATCAATGTCCACGACGTGTGAACAAGTTCAATCAATTA | 1080 |
| O | y | 1152 | AACGAGACAGCTCTCAGATTTAACTGACGATCTCAGACGTCAGCTGACTCATGTGC | 1211 |
| D | b | 1081 | AACGAGACAGCTCTCAGATTTAACTGACGATCTCAGACGTCAGCTGACTCATGTGC | 1140 |
| O | y | 1212 | ATTTCCTTTCTCGGCCAGTCTGGGGCTGGGGTGTCAAGCTGGGGACATCGGCGCTGGCT | 1272 |
| D | b | 1141 | ATTTCCTTTCTCGGCCAGTCTGGGGCTGGGGTGTCAAGCTGGGGACATCGGCGCTGGCT | 1200 |
| O | y | 1272 | GCTGGTCTGTGTTCTGTGTGGCTGGCATGTGTCTATCTCATTTGCTTGGCTGTCTGA | 1331 |
| D | b | 1201 | GCTGGTCTGTGTTCTGTGTGGCTGGCATGTGTCTCTATCTCATTTGCTTGGCTGTCTGA | 1260 |

OY 1332 GTGCGCGGAAAGAACTACGCGGAGCTGACATCTTTCCAGCCCGGGATCTACCATCC 1391
DB 1361 GTGCGCGGAAAGAACTACGCGGAGCTGACATCTTTCCAGCCCGGGATCTACCATCC 1320
OY 1392 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1451
DB 1321 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1380
OY 1452 TCGTAGCCCTTATGAGAGTCTGAGAGTATGAGTGGACAGACCTCTTACACAAA 1511
DB 1381 TCGTAGCCCTTATGAGAGTCTGAGAGTATGAGTGGACAGACCTCTTACACAAA 1440
OY 1512 CCGACGAGGAGGAGGACCTTGTGCACTTGTAGGGGACAGCGCCCTCTGAGCTAGTG 1571
DB 1441 CCGACGAGGAGGAGGACCTTGTGCACTTGTAGGGGACAGCGCCCTCTGAGCTAGTG 1500
OY 1572 G 1572
DB 1501 G 1501

RESULT 9

US-60-453-135-343
; Sequence 343, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001456
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/60/453,135
; NUMBER OF SEQ ID NOS: 2003-03-10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 343
; LENGTH: 1549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-453-135-343

Query Match 89.7%, Score 1409.8; DB 11; Length 1549;

Best Local Similarity 95.1%; Pred. No. 2.3e-264;

Matches 148; Conservative 5; Mismatches 12; Indels 60; Gaps 1;

OY 12 CTGCTGAATCTGCTGCGCCCTCCCACTTTCACACACACACACACACACACACAC 71
DB 1 CTGCTGAATCTGCTGCGCCCTCCCACTTTCACACACACACACACACACACACAC 60
OY 72 CCACTCTCTTCTCTGCTGCTGCTCTCTACAGTCTTACAGTTTACAGGTTCTGG 131
DB 61 CCACTCTCTTCTCTGCTGCTGCTCTCTACAGTCTTACAGTTTACAGGTTCTGG 120
OY 132 TCATGCAAGCTTACCCAGGTGAGAAAGAGACTTGGCTACCCAGAGAGTTCACT 191
DB 121 TCATGCAAGCTTACCCAGGTGAGAAAGAGACTTGGCTACCCAGAGAGTTCACT 180
OY 192 GCCAGCTTACGAGAAAGATGCTGAGATGACACACACACACACACACACACAC 251
DB 181 GCCAGCTTACGAGAAAGATGCTGAGATGACACACACACACACACACACACAC 240
OY 252 CCGCGGTTTCAAGGCTCTCTCCACCTCAGAGGACAGATGCTACTGGCCCGGACACGA 311
DB 241 CCGCGGTTTCAAGGCTCTCTCCACCTCAGAGGACAGATGCTACTGGCCCGGACACGA 300
OY 312 ACCAGCTTCAAGGTTTCAAGGCTCTCTCCACCTCAGAGGACAGATGCTACTGGCCCGGACACAG 371
DB 301 ACCAGCTTCAAGGTTTCAAGGCTCTCTCCACCTCAGAGGACAGATGCTACTGGCCCGGACACAG 360
OY 372 GCCAGCCCTGGGCTCCACACACCCCGACACACACATGCTACCTCAGCCCGGACACAA 431
DB 361 GCCAGCCCTGGGCTCCACACACCCCGACACACACATGCTACCTCAGCCCGGACACAA 420

OY 432 GCCAGCCCGGCTCCACACCCCGCCCGCCACAGGTGTCACCTCGGCCCGGACACACAG 491
DB 421 GCCAGCCCGGCTCCACACCCCGCCCGCCACAGGTGTCACCTCGGCCCGGACACACAG 448
OY 492 GCCCGCCCCGGGCTCACCGCCCGCGCCCGGAGTGTGACCTCGGCCCGGACACACAG 551
DB 449 -----GCCAGGTTGTCACTCGGCCCGGAGTGTGACCTCGGCCCGGACACACAG 480
OY 552 GCCCGCCCCGGGCTCCACACCCCGCCCGCCCGGAGTGTGACCTCGGCCCGGACACACAG 611
DB 481 GCCCGCCCCGGGCTCCACACCCCGCCCGCCCGGAGTGTGACCTCGGCCCGGACACACAG 540
OY 612 GCCCGCTTGGGCTCCACACCCCGCCCGCCCGGAGTGTGACCTCGGCCCGGAGTGTGAC 671
DB 541 GCCCGCTTGGGCTCCACACCCCGCCCGCCCGGAGTGTGACCTCGGCCCGGAGTGTGAC 600
OY 672 ATCAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTCTGCGAGGCTTACACAAACCC 731
DB 601 ATCAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTCTGCGAGGCTTACACAAACCC 660
OY 732 AGCAGCAAGAGCACTCCACCCAGACATTCACCCACCACTGTATATCTTACACACCT 791
DB 661 AGCAGCAAGAGCACTCCACCCAGACATTCACCCACCACTGTATATCTTACACACCT 720
OY 792 TGCAGACCATGAGCAACAGACATGATGCGATGACATCACCATGACAGGATCTCTCT 851
DB 721 TGCAGACCATGAGCAACAGACATGATGCGATGACATCACCATGACAGGATCTCTCT 780
OY 852 CACCTCTTCAATCAGACACTTCTCCCACTGTTGTACTGAGGCTCTCTTCTTCTTCT 911
DB 781 CACCTCTTCAATCAGACACTTCTCCCACTGTTGTACTGAGGCTCTCTTCTTCTTCT 840
OY 912 GTCTTTTCAATTTCAAACTCTCAGTTTAATCTCTCTGTGAGATCCAGACACGACTA 971
DB 841 GTCTTTTCAATTTCAAACTCTCAGTTTAATCTCTGTGAGATCCAGACACGACTA 900
OY 972 CTACCAAGGCTGAGAGAGACATTTCTGAATGTTTTTGCAGATTATTAACAAGGGGG 1031
DB 901 CTACCAAGGCTGAGAGAGACATTTCTGAATGTTTTTGCAGATTATTAACAAGGGGG 960
OY 1032 TTTTCTGGGCTCTTCAATTTAAGTTACAGCCAGGATCTGTGTGTATGATGACTCT 1091
DB 961 TTTTCTGGGCTCTTCAATTTAAGTTACAGCCAGGATCTGTGTGTATGATGACTCT 1020
OY 1092 GGCCTTCCAGAAAGGTACATCATATGTCCACAGACGTGAGACACAGTTCAATGATTA 1151
DB 1021 GGCCTTCCAGAAAGGTACATCATATGTCCACAGACGTGAGACACAGTTCAATGATTA 1080
OY 1152 AACGAAAGACGCTTGTATATTAACCTGACATCTTCAAGCTCAGGCTGATGATGCTC 1211
DB 1081 AACGAAAGACGCTTGTATATTAACCTGACATCTTCAAGCTCAGGCTGATGATGCTC 1140
OY 1212 ATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGACAGCTGAGGACATCGCGCTGTGT 1271
DB 1141 ATTTCTTTCTCTGCGCAGTCTGAGGCTGAGGCTGACAGCTGAGGACATCGCGCTGTGT 1200
OY 1272 GCTGT 1331
DB 1201 GCTGT 1260
OY 1332 GTGCGCGGAAAGAACTACGCGGAGCTGACATCTTTCCACCCCGGATCTTACACATCC 1391
DB 1261 GTGCGCGGAAAGAACTACGCGGAGCTGACATCTTTCCACCCCGGATCTTACACATCC 1320
OY 1392 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1451
DB 1321 TATGAGCGAGTACCCCACTTACACACACACACACACACACACACACACACACAC 1380
OY 1452 TCGTAGCCCTTATGAGAGTCTGAGAGTATGAGTGGACAGACCTCTTACACAAA 1511
DB 1381 TCGTAGCCCTTATGAGAGTCTGAGAGTATGAGTGGACAGACCTCTTACACAAA 1440
OY 1512 CCGACGAGGAGGAGGACCTTGTGCACTTGTAGGGGACAGCGCCCTCTGAGCTAGTG 1571


```

; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038PC
; CURRENT APPLICATION NUMBER: PCT/US02/19669A
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 310
; LENGTH: 4139
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-19669A-310

Query Match      75.7%; Score 1190.2; DB 1; Length 4139;
Best Local Similarity 94.8%; Pred. No. 1.2e-221;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Oy 249 CAGCCCGGTTGAGCTCTCTCCACACTCAGGAGCAGATGTCACCTGCGCCCGGCAC 308
Db 2545 CAGCCCGGTTGAGCTCTCTCCACACTCAGGAGCAGATGTCACCTGCGCCCGGCAC 2604
Oy 309 GGAACAGAGTTGAGTTGAGTGCACCTGAGGAGAGATGTCACCTGAGTCCAGTTCAC 368
Db 2605 CAGGCGGCGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 2664
Oy 369 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 428
Db 2665 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 2724
Oy 429 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 488
Db 2725 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 2784
Oy 489 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 548
Db 2785 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 2844
Oy 549 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 608
Db 2845 CAGGCGGCGCGGCTCCACCGCCCGCCAGCCACGAGTGTCACTTGCGCCCGGACAC 2904
Oy 609 CAGGCGGCGCTTGAGCTCCACCGCCCTCCAGTGCACATGTCACCTTGAGCTCAGGCTC 668
Db 2905 CAGGCGGCGCTTGAGCTCCACCGCCCTCCAGTGCACATGTCACCTTGAGCTCAGGCTC 2964
Oy 669 TGCATCAGGCTCAGCTTCTACTCTGTGTGCAACAGGCACTCTGCGAGGCTACACAC 728
Db 2965 TGCATCAGGCTCAGCTTCTACTCTGTGTGCAACAGGCACTCTGCGAGGCTACACAC 3024
Oy 729 CCGAGCCAGAGAGACTCACAACCGAGATTCCAGCCACACCTGTATATCTCTACAC 788
Db 3025 CCGAGCCAGAGAGACTCACAACCGAGATTCCAGCCACACCTGTATATCTCTACAC 3084
Oy 789 CCTTGCCAGGACATGACCAAGATGTCAGTACATGTCACCTGACCTGAGTACCTCC 848
Db 3085 CCTTGCCAGGACATGACCAAGATGTCAGTACATGTCACCTGACCTGAGTACCTCC 3144
Oy 849 TCTCAGCTCTCTCATATCAGACACTTCTCCCGAGTGTCTATATGAGGCTCTTTTCTTTT 908
Db 3145 TCTCAGCTCTCTCATATCAGACACTTCTCCCGAGTGTCTATATGAGGCTCTTTTCTTTT 3204
Oy 909 CCTGCTTTTTCACATTGAAACCTCCAGTTTAATCTCTCTGAGAGATCCAGACCGA 968

```

```

Db 3205 CCTGCTTTTTCACATTGAAACCTCCAGTTTAATCTCTCTGAGAGATCCAGACCGA 3264
Oy 969 CTACTCTCAAGAGCTGCAGAGAGACATTTCTGAATATGTTTGGATTTATTAACAAG 1028
Db 3265 CTACTCTCAAGAGCTGCAGAGAGACATTTCTGAATATGTTTGGATTTATTAACAAG 3324
Oy 1029 GGGTTTCTGAGGCTCTCCAAATATTAAGTTACAGCCAGAGATCTGTGTGATCAATTGAC 1088
Db 3325 GGGTTTCTGAGGCTCTCCAAATATTAAGTTACAGCCAGAGATCTGTGTGATCAATTGAC 3384
Oy 1089 TCTGAGCTTCCAGAGAGTACCATCATATGTCCACAGCTGAGAGACAGATTCATCAGTA 1148
Db 3385 TCTGAGCTTCCAGAGAGTACCATCATATGTCCACAGCTGAGAGACAGATTCATCAGTA 3444
Oy 1149 TAAACGAGAGAGCTCTTGATATTAACCTGACATCTGACATCTGAGCTGATCAGT 1208
Db 3445 TAAACGAGAGAGCTCTTGATATTAACCTGACATCTGACATCTGAGCTGATCAGT 3504
Oy 1209 GCCATTTCCTTCTCTGCCCAGTCTGAGGCTGAGGCTGACAGCTGAGGCAATGCGCTGCT 1268
Db 3505 GCCATTTCCTTCTCTGCCCAGTCTGAGGCTGAGGCTGACAGCTGAGGCAATGCGCTGCT 3564
Oy 1269 GGTGCTGATCTGTGTTGCTGAGTGGCGCTGAGCATTGTCTATCTCATTTGCGCTGTG 1328
Db 3565 GGTGCTGATCTGTGTTGCTGAGTGGCGCTGAGCATTGTCTATCTCATTTGCGCTGTG 3624
Oy 1329 TCAGTGCGCGGAAAGAACTACGAGGAGCTGAGACATCTTTCAGCCCGGATACCTACA 1388
Db 3625 TCAGTGCGCGGAAAGAACTACGAGGAGCTGAGACATCTTTCAGCCCGGATACCTACA 3684
Oy 1389 TCCATGACGAGTACCCCACTACACACACCCATGAGGCGCTATGTCCTTACAGTAC 1448
Db 3685 TCCATGACGAGTACCCCACTACACACACCCATGAGGCGCTATGTCCTTACAGTAC 3744
Oy 1449 CGATGTCACCCCTATAGAGAGTTCTGACATTAATGTGAGGAGCAGCCTCTTACAC 1508
Db 3745 CGATGTCACCCCTATAGAGAGTTCTGACATTAATGTGAGGAGCAGCCTCTTACAC 3804
Oy 1509 AAACCCAGAGTGGAGGCACTTCTSCCAACTTTGAGGG 1547
Db 3805 AAACCCAGAGTGGAGGCACTTCTSCCAACTTTGAGGG 3843

RESULT 12
US-10-170-235-24364
; Sequence 24364, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; FILE REFERENCE: C1001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 24364
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-24364

Query Match      60.7%; Score 953.8; DB 8; Length 1353;
Best Local Similarity 81.0%; Pred. No. 8.1e-176;
Matches 1266; Conservative 0; Mismatches 12; Indels 285; Gaps 2;

Oy 10 GGGTGTGTAATCTGTTCTGCGCCCTCCGACCACTTTCAACACACACATGACACCGGGC 69
Db 19 GCTGCTGAATCTGTTCTGCGCCCTCCGACCACTTTCAACACACACATGACACCGGGC 78
Oy 70 ACCGAGTCTCTTCTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTTACAGGTTCT 129
Db 79 ACCGAGTCTCTTCTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTTACAGGTTCT 138

```

```
QY 130 GGTATGCAAGCTTACACCCAGGTGGAGAAAAGAGACTTCGGCTACCCAGAGAGTTCA 189
Db 139 GGTATGCAAGCTTACACCCAGGTGGAGAAAAGAGACTTCGGCTACCCAGAGAGTTCA 198
QY 190 GTGCCAGCTTACTAGAGAAATGCTGTAGATGACACAGCGCTACTCTCCAGCCAC 249
Db 199 GTGCCAGCTTACTAGAGAAATGCTGTAGATGACACAGCGCTACTCTCCAGCCAC 258
QY 250 AGCCCCGGTTCAAGGCTCTCCACCACTCAGGAGACGAGATGTCACTCTGCCCCGAGCAG 309
Db 259 AGCCCCGGTTCAAGGCTCTCCACCACTCAGGAGACGAGATGTCACTCTGCCCCGAGCAG 318
QY 310 GAACAGCTTACAGGTTCAAGTGCACCTGGGAGACAGAGATGTCACTCTGCCCCGAGCAG 369
Db 319 GAACAGCTTACAGGTTCAAGTGCACCTGGGAGACAGAGATGTCACTCTGCCCCGAGCAG 378
QY 370 AGGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 429
Db 379 AGGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 438
QY 430 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 489
Db 439 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 498
QY 490 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 549
Db 497 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 558
QY 550 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 609
Db 559 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 618
QY 610 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 669
Db 619 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 678
QY 670 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 729
Db 679 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 738
QY 730 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 789
Db 739 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 798
QY 790 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 849
Db 799 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 858
QY 850 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 909
Db 859 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 918
QY 910 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 969
Db 919 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 978
QY 970 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1029
Db 979 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1038
QY 1030 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1089
Db 1039 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1098
QY 1090 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1149
Db 1099 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1158
QY 1150 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1209
Db 1159 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1218
QY 1210 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1269
Db 1219 AAGCCAGCTTCTGAGCTTCAACACCCCGCAGCCAGATGTCACTCTGAGCCCGAGCAG 1278
```

```
Db 954 -----GTCGGGGCTGGGGTCCAGGCTGGGGCATGGGCTGGT 993
QY 1270 GTGCTGATCTGATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1329
Db 994 GTGCTGATCTGATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1053
QY 1330 CAGTGGCCGGAAGAACTACAGGAGCAGTGGACATCTTTCCAGCCGGAGATACCTACAT 1389
Db 1054 CAGTGGCCGGAAGAACTACAGGAGCAGTGGACATCTTTCCAGCCGGAGATACCTACAT 1113
QY 1390 CCTATGAGCAGATACCCCACTACACCACTACACCACTACACCACTACACCACTACACCACT 1449
Db 1114 CCTATGAGCAGATACCCCACTACACCACTACACCACTACACCACTACACCACTACACCACT 1173
QY 1450 GATGTGATCTCTGATGAGAGATCTTCTGAGATGATGATGATGATGATGATGATGATGAT 1509
Db 1174 GATGTGATCTCTGATGAGAGATCTTCTGAGATGATGATGATGATGATGATGATGATGAT 1233
QY 1510 AACCAGCAGTGGCAGCAGCAGCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1569
Db 1234 AACCAGCAGTGGCAGCAGCAGCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1293
QY 1570 TGG 1572
Db 1294 TGG 1296

RESULT 13
US-60-452-680-474
; Sequence 474, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLOU1450
; CURRENT APPLICATION NUMBER: US/60/452,680
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 474
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-474

Query Match 60.6%; Score 952.2; DB 11; Length 1353;
Best Local Similarity 80.7%; Pred. No. 1.6e-175;
Matches 1262; Conservative 4; Mismatches 12; Indels 285; Gaps 2;

QY 10 GGTGCTTGAATCTGCTTGTGCCCCCTCCCACTTTCACACCATGACACCGGAGC 69
Db 19 GGTGCTTGAATCTGCTTGTGCCCCCTCCCACTTTCACACCATGACACCGGAGC 78
QY 70 ACCAGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 129
Db 79 ACCAGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 138
QY 130 GGTGATCAAGCTTACCCAGGTGAGAGAAAGAGAGACTTGGCTACCCAGAGAGTTCA 189
Db 139 GGTGATCAAGCTTACCCAGGTGAGAGAAAGAGAGACTTGGCTACCCAGAGAGTTCA 198
QY 190 GTGCCAGCTTACTAGAGAAATGCTGTAGATGACACAGCGCTACTCTCCAGCCAC 249
Db 199 GTGCCAGCTTACTAGAGAAATGCTGTAGATGACACAGCGCTACTCTCCAGCCAC 258
QY 250 AGCCCCGGTTCAAGGCTCTCCACCACTCAGGAGACGAGATGTCACTCTGCCCCGAGCAG 309
Db 259 AGCCCCGGTTCAAGGCTCTCCACCACTCAGGAGACGAGATGTCACTCTGCCCCGAGCAG 318
QY 310 GAACAGCTTACAGGTTCAAGTGCACCTGGGAGACAGAGATGTCACTCTGCCCCGAGCAG 369
Db 319 GAACAGCTTACAGGTTCAAGTGCACCTGGGAGACAGAGATGTCACTCTGCCCCGAGCAG 378
```

```

Db 319 GAACCAAGCTTCAAGTTCACTGCTCCACCTGGGGACAGAAATGTCACCTGGTCCAGTACCC 378
Qy 370 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 429
Db 379 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 438
Qy 430 AAGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 489
Db 439 AAGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 466
Qy 490 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 549
Db 467 -----CAGCCCAAGTGTACCTCAGCCCGGACAA 498
Qy 550 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 609
Db 499 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 558
Qy 610 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 669
Db 559 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 618
Qy 670 GCATCAGGCTCAGCTTCTACTCTGTGTGCAACAGGACCTCTGCAGGAGCTACCAAC 729
Db 619 GCATCAGGCTCAGCTTCTACTCTGTGTGCAACAGGACCTCTGCAGGAGCTACCAAC 678
Qy 730 CGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
Db 679 CGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
Qy 790 CTTGCGAGCCATAGACCAAGAGCTGATGCAAGTACCTCAGCAGTACAGGAGAGAGAG 849
Db 739 CTTGCGAGCCATAGACCAAGAGCTGATGCAAGTACCTCAGCAGTACAGGAGAGAGAG 798
Qy 850 CTGACCTCTCCATCAGAGAGCTTCTCCAGAGTGTCTACTGAGGAGTCTCTTCTTTTC 909
Db 799 CTGACCTCTCCATCAGAGAGCTTCTCCAGAGTGTCTACTGAGGAGTCTCTTCTTTTC 858
Qy 910 CTGCTTTTCAATTTCAACCTCAGATTAATCTCTCTGGAAGATCCGAGACCGAG 969
Db 859 CTGCTTTTCAATTTCAACCTCAGATTAATCTCTCTGGAAGATCCGAGACCGAG 918
Qy 970 TACTACCAAGAGCTGAGAGAGATTTCTGAATGTTTGGCAGATTTTAAACAAGAG 1029
Db 919 TACTACCAAGAGCTGAGAGAGATTTCTGAATGTTTGGCAGATTTTAAACAAGAG 953
Qy 1030 GGTTCCTGGGCTCTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1089
Db 954 ----- 953
Qy 1090 CTGAGCTTCCGAGAGAGTACATCATGTCCAGAGCTGGAGACAAGTTCAATCAGTAT 1149
Db 954 ----- 953
Qy 1150 AAAAGGAGAGAGCTCTGATATATCTGAGAGTCTGAGAGTGAAGTATGTTG 1209
Db 954 ----- 953
Qy 1210 CCAATTTCTTTCTGCGCAAGTCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGG 1269
Db 954 -----GTCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGG 993
Qy 1270 GTGCTGGTCTGTTCTGTTGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGG 1329
Db 994 GTGCTGGTCTGTTCTGTTGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGG 1053
Qy 1330 CAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1389
Db 1054 CAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
Qy 1390 CTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1449
Db 1114 CTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173

```

```

Qy 1450 GATCTGAGCCCTTATGAGAGTTTCTGCAAGTAAATGTTGGACAGACCTCTCTTACACA 1509
Db 1174 GATCTGAGCCCTTATGAGAGTTTCTGCAAGTAAATGTTGGACAGACCTCTCTTACACA 1233
Qy 1510 AAGCGAGGCTGAGAGCACTTCTGCAACTTGTAGGGGACAGTGGCCTTGAAGTGA 1569
Db 1234 AAGCGAGGCTGAGAGCACTTCTGCAACTTGTAGGGGACAGTGGCCTTGAAGTGA 1293
Qy 1570 TGG 1572
Db 1294 TGG 1296

RESULT 14
us-60-453-135-346
; Sequence 346, Application us/60453135
; GENERAL INFORMATION:
; APPLICANT: CARBELL, Michele
; APPLICANT: IAKUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001456
; CURRENT APPLICATION NUMBER: us/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 346
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Homo sapiens
us-60-453-135-346

Query Match 60.6%; Score 952.2; DB 11; Length 1353;
Best Local Similarity 80.7%; Pred. No. 1,6e-175;
Matches 1262; Conservative 4; Mismatches 12; Indels 285; Gaps 2;

Qy 10 GGTGTTGTAATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 69
Db 19 GCTGCTGTAATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 78
Qy 70 ACCAGTCTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129
Db 79 ACCAGTCTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138
Qy 130 GGTATGAGAGCTTACCCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
Db 139 GGTATGAGAGCTTACCCAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
Qy 190 GTGCCAGCTTACTGAGAGAGTCTGTGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 249
Db 199 GTGCCAGCTTACTGAGAGAGTCTGTGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 258
Qy 250 AGCCCGGTTGAGGCTCTCTCAACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Db 259 AGCCCGGTTGAGGCTCTCTCAACCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
Qy 310 GAACAGCTTCAAGTTCAAGTGTGCACTGGGAGACAGATGTACCTCGGTCCAGTACAC 369
Db 319 GAACAGCTTCAAGTTCAAGTGTGCACTGGGAGACAGATGTACCTCGGTCCAGTACAC 378
Qy 370 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 429
Db 379 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 438
Qy 430 AAGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 489
Db 439 AAGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 466
Qy 490 AGGCGAGCCTTGGGCTCCACACCCCGGACAGCCCAAGTGTACCTCAGCCCGGACAA 549
Db 467 -----CAGCCCAAGTGTACCTCAGCCCGGACAA 498

```


QY 550 AGGCGGAGCCCGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 609
DB 499 AGGCGGAGCCCGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 558
QY 610 AGGCGGAGCTTGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 669
DB 559 AGGCGGAGCTTGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 618
QY 670 GCATCAGGCTCAGCTTCTACTCTGAGTCAACCGGAGCTCTGCGGAGGAGTCAACCAAC 729
DB 619 GCATCAGGCTCAGCTTCTACTCTGAGTCAACCGGAGCTCTGCGGAGGAGTCAACCAAC 678
QY 730 CCAAGCAGAGAGAGAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGC 789
DB 679 CCAAGCAGAGAGAGAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGC 738
QY 790 CTTGCGAGCAGTACAGCAGCAGTATGAGTACAGTACAGTACAGTACAGTACAGTACAGT 849
DB 739 CTTGCGAGCAGTACAGCAGCAGTATGAGTACAGTACAGTACAGTACAGTACAGTACAGT 798
QY 850 CTCACCTCTCCAGTACAGCAGCTTCTCCAGGAGTCTACTGAGGAGTCTCTCTCTCTCTCT 909
DB 799 CTCACCTCTCCAGTACAGCAGCTTCTCCAGGAGTCTACTGAGGAGTCTCTCTCTCTCTCT 858
QY 910 CTGCTCTTTCATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 969
DB 859 CTGCTCTTTCATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 918
QY 970 TACTACCAAGAGCTGAGAGAGCATTTCTGAATTTGTTTTCAGATTATTAACAAGAG 1029
DB 919 TACTACCAAGAGCTGAGAGAGCATTTCTGAATTTGTTTTCAGATTATTAACAAGAG 953
QY 1030 GGTCTTTCGAGGCTCTCCAGTATTAAGTTCAAGGAGCAGGAGTCTGTGTGTACATTTGACT 1089
DB 954 ----- 953
QY 1090 CTGCGCTTCGAGAGAGTACATCAATGTCCAGCAGTGGAGACAGATTCAATCAGATTAT 1149
DB 954 ----- 953
QY 1150 AAAACGGAAGAGCTCTCGATATTAACCTGAGATCTCAGAGCTCAGGAGTCAATGTG 1209
DB 954 ----- 953
QY 1210 CCATTTCT 1269
DB 954 ----- 953
QY 1270 GTGCTGCTCTGCT 1329
DB 994 GTGCTGCTCTGCT 1053
QY 1330 CAGTCCGCGGAGAGAGTCAAGGAGTCAAGTCAATCTTTCAGAGCAGGAGTCAAGTCAAT 1389
DB 1054 CAGTCCGCGGAGAGAGTCAAGGAGTCAAGTCAATCTTTCAGAGCAGGAGTCAAGTCAAT 1113
QY 1390 CCTATGAGCAGTACCCAGTCAACCACTGAGGAGTATGAGTCTGAGTCTGAGTCTGAGTCT 1449
DB 1114 CCTATGAGCAGTACCCAGTCAACCACTGAGGAGTATGAGTCTGAGTCTGAGTCTGAGTCT 1173
QY 1450 GATCGTAGCCCTATGAGAGGTTCTGAGAGTAAAGTGGAGAGAGCTCTCTCTCTCTCTCT 1509
DB 1174 GATCGTAGCCCTATGAGAGGTTCTGAGAGTAAAGTGGAGAGAGCTCTCTCTCTCTCTCT 1233
QY 1510 AACCCAGAGTGGAGAGCT 1569
DB 1234 AACCCAGAGTGGAGAGCT 1293
QY 1570 TGG 1572
DB 1294 TGG 1296

RESULT 15
US-60-453-050-346
Sequence 346, Application US/60453050
GENERAL INFORMATION:
APPLICANT: CARILL, Michele
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001457
CURRENT APPLICATION NUMBER: US/60/453, 050
NUMBER OF SEQ ID NOS: 82762
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 346
LENGTH: 1353
TYPE: DNA
ORGANISM: Homo sapiens
US-60-453-050-346
Query Match 60.6%; Score 952.2; DB 11; Length 1353;
Best Local Similarity 80.7%; Pred. No. 1.6e-175;
Matches 1262; Conservative 4; Mismatches 12; Indels 285; Gaps 2;
QY 10 GGTGCTGTAATCTGTTCTGCGCCCTCTCCCACTTTTCAACCAACCAACCAACCAACCAAC 69
DB 19 GGTGCTGTAATCTGTTCTGCGCCCTCTCCCACTTTTCAACCAACCAACCAACCAACCAAC 78
QY 70 ACCAGTCTCTTCT 129
DB 79 ACCAGTCTCTTCT 138
QY 130 GGTGAGCAAGCTCTTACCCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
DB 139 GGTGAGCAAGCTCTTACCCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
QY 190 GTGCGCAGCTCTTACAG 249
DB 199 GTGCGCAGCTCTTACAG 258
QY 250 AGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
DB 259 AGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
QY 310 GAACCAAGCTTCAAGTTCAGCTGCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
DB 319 GAACCAAGCTTCAAGTTCAGCTGCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 378
QY 370 AGGCAAGCTTCAAGTTCAGCTGCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
DB 379 AGGCAAGCTTCAAGTTCAGCTGCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
QY 430 AAGCAGAGCCCGGAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 489
DB 439 AAGCAGAGCCCGGAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 466
QY 490 AAGCAGAGCCCGGAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 549
DB 467 ----- 466
QY 550 AGGCGGAGCCCGGAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 609
DB 499 AGGCGGAGCCCGGAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 558
QY 610 AGGCGGAGCTTGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 669
DB 559 AGGCGGAGCTTGGAGCTCCACCGCCCGGCGGAGGAGTCACTCGGCGCCCGGAGAAC 618
QY 670 GCATCAGGCTCAGCTTCTACTCTGAGTCAACCGGAGCTCTGCGGAGGAGTCAACCAAC 729
DB 619 GCATCAGGCTCAGCTTCTACTCTGAGTCAACCGGAGCTCTGCGGAGGAGTCAACCAAC 678
QY 730 CCAAGCAGAGAGAGAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGCAGCTCCAGCAGC 789


```

Db 679 CCAGCCAGCAGAGCATCTCAATTCATTTCCAGCCACCACTGTATATCTTACACC 738
QY 790 CTTGCGAGCCATAGCAGCAAGACTGATGCCAGTAGACCTCACCATAGCAGGTAACCTCT 849
Db 739 CTTGCCAGCCATAGCAGCAAGACTGATGCCAGTAGACCTCACCATAGCAGGTAACCTCT 798
QY 850 CTGACCTCTCAATCAAGCAGCATTTCTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC 909
Db 799 CTCACCTCTCCCAATCAAGCAGCATTTCTCCCAAGTGTCTACTGCGGCTCTCTTTCTTTTC 858
QY 910 CTGCTTTTCAATTTCAAACTCAGTTTAATTTCTCTCTGGAAGATCCAGCAGCCAGC 969
Db 859 CTGCTTTTCAATTTCAAACTCAGTTTAATTTCTCTCTGGAAGATCCAGCAGCCAGC 918
QY 970 TACTACCAAGAGCTGAGAGAGACATTTTGAATGTTTTTGCAATTTATTAACAAGG 1029
Db 919 TACTACCAAGAGCTGAGAGAGACATTTTGAATGTTTTTGCAATTTATTAACAAGG 953
QY 1030 GGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAGAGATCTGTGGTGGTACAAATGACT 1089
Db 954 ----- 953
QY 1090 CTGGCTTCGAGAGAGTACCATCATATGTCACGACGTGAGACACAGTTCAATCAGTAT 1149
Db 954 ----- 953
QY 1150 AAAACGAGAGCAGCCTCTGATATATACTGACATCTCAGACGTACGCGTAGTCAATGTG 1209
Db 954 ----- 953
QY 1210 CCATTTCTTTCTGCGCCAGTCTGCGGCTGCGGCTGCGAGCTGCGGCAATGCGCTGCTG 1269
Db 954 -----GTCGCGGCTGCGGCTGCGAGCTGCGGCAATGCGCTGCTG 993
QY 1270 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Db 994 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1053
QY 1330 CAGTGCCTGCGAAGAACTACGCGGAGCTGAGCATTTTCCAGCCCGGAGTACCTACCAT 1389
Db 1054 CAGTGCCTGCGAAGAACTACGCGGAGCTGAGCATTTTCCAGCCCGGAGTACCTACCAT 1113
QY 1390 CCTATGAGAGTACCCGACCTACGACCCGATGCGGCTATGTCCTTACGAGTACC 1449
Db 1114 CCTATGAGAGTACCCGACCTACGACCCGATGCGGCTATGTCCTTACGAGTACC 1173
QY 1450 GATGTAGCCCTATGAGAGGTTTCTGAGGTATGTTGAGAGAGCTCTCTTTACACA 1509
Db 1174 GATGTAGCCCTATGAGAGGTTTCTGAGGTATGTTGAGAGAGCTCTCTTTACACA 1233
QY 1510 AACCCAGAGTGGCAGCCTCTGCAACTTGTAGGGGAGCAGTGGCCTCTGAGCTGAG 1569
Db 1234 AACCCAGAGTGGCAGCCTCTGCAACTTGTAGGGGAGCAGTGGCCTCTGAGCTGAG 1293
QY 1570 TGG 1572
Db 1294 TGG 1296

```

Search completed: May 8, 2003, 18:14:50
 Job time : 797 secs


```
Db 121 ACAGGTTCTGGGTCAATGCAAGCTCTACCCAGGTGAGAAAAGAGACTTCGGCTCCAG 180
Qy 181 AGAAGTTCTAGTCCCAAGCTCTAAGAGAAATGTGTGATGTGACCAAGCAAGCTACTC 240
Db 181 AGAAGTTCTAGTCCCAAGCTCTAAGAGAAATGTGTGATGTGACCAAGCAAGCTACTC 240
Qy 241 TCCAGCAACAGCCCGGTTGAGGCTCTCAACATCGAGGAGAGAGATGTCACTTGGCC 300
Db 241 TCCAGCAACAGCCCGGTTGAGGCTCTCAACATCGAGGAGAGAGATGTCACTTGGCC 300
Qy 301 CCGGCAACGAGAAACAGCTTCAAGTTCACTGACCTGAGGAGACAGATGTCACTTGGCT 360
Db 301 CCGGCAACGAGAAACAGCTTCAAGTTCACTGACCTGAGGAGACAGATGTCACTTGGCT 360
Qy 361 CCAATCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 420
Db 361 CCAATCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 420
Qy 421 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 480
Db 421 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 480
Qy 481 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 540
Db 481 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 540
Qy 514 -----CCGCGCGCCACCGGTGTCACCTCGGCC 540
Db 514 -----CCGCGCGCCACCGGTGTCACCTCGGCC 540
Qy 541 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 600
Db 541 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 600
Qy 601 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 660
Db 601 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 660
Qy 661 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 720
Db 661 CCGGCAACAGGCAAGCCCTGAGGCTCAACCAACCCGAGCCAGCAAGATGTCACTCAGCC 720
Qy 721 TCGAGCTGTGATCAAGGCTCAAGCTTCTGATGTGACCAACGAGCACTTCCAGGCT 780
Db 721 TCGAGCTGTGATCAAGGCTCAAGCTTCTGATGTGACCAACGAGCACTTCCAGGCT 780
Qy 781 ACCAACAACCCAGGCAAGCACTTCCAGGCAAGCACTTCCAGGCAAGCACTTCCAGGCT 840
Db 781 ACCAACAACCCAGGCAAGCACTTCCAGGCAAGCACTTCCAGGCAAGCACTTCCAGGCT 840
Qy 841 GTAAGCTGTGATCAAGGCTCAAGCTTCCAGGCAAGCACTTCCAGGCTGTCTCT 900
Db 841 GTAAGCTGTGATCAAGGCTCAAGCTTCCAGGCAAGCACTTCCAGGCTGTCTCTCT 900
Qy 901 TTTCTTTCTGTCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 960
Db 901 TTTCTTTCTGTCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 960
Qy 961 AGAAGCAAGCTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1020
Db 961 AGAAGCAAGCTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1020
Qy 1021 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1081 CAATTTGCTGTGCTTCCGAGAGAGTCAATCAATGTGACAGAGCTGTGAGAGACAGTTT 1140
Db 1141 CAATTTGCTGTGCTTCCGAGAGAGTCAATCAATGTGACAGAGCTGTGAGAGACAGTTT 1140
Qy 1141 AATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 AATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
```

```
Db 1201 AATCAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1201 AGTCAATGCAATTTCTTTTCTGTGACCAAGTGTGGGGCTGGGGGCTGGGGGCTGGGGGCTG 1260
Db 1261 AGTCAATGCAATTTCTTTTCTGTGACCAAGTGTGGGGGCTGGGGGCTGGGGGCTGGGGGCTG 1320
Qy 1261 GCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
Db 1321 GCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Qy 1321 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
Db 1381 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
Qy 1381 ACTTACATCTATGAGGAGAGTACCCCACTACCAACCCAGTGGGCTATGTGCCCCCT 1440
Db 1441 ACTTACATCTATGAGGAGAGTACCCCACTACCAACCCAGTGGGCTATGTGCCCCCT 1500
Qy 1441 AGCAGTACCGATGTGAGCCCTTATGAGAGAGTTCTGAGAGTAAATGTGGAGAGAGCTC 1500
Db 1501 AGCAGTACCGATGTGAGCCCTTATGAGAGAGTTCTGAGAGTAAATGTGGAGAGAGCTC 1560
Qy 1501 TCTTACAAACCCAGAGTGGAGAGCACTTCTGCAACTTGTAGGGGCAAGTGGCCCTC 1560
Db 1561 TCTTACAAACCCAGAGTGGAGAGCACTTCTGCAACTTGTAGGGGCAAGTGGCCCTC 1620
Qy 1561 TGAAGTGTAGTGG 1572
Db 1621 TGAAGTGTAGTGG 1632

RESULT 2
US-09-967-768A-224
Sequence 224: Application US/09967768A
Patent No. US2002150877A1
GENERAL INFORMATION:
APPLICANT: Augmentus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-72
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236,109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,111
NUMBER OF SEQ ID NOS: 325
SOFTWARE: PatentIn version 3.0
SEQ ID NO 224
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-224

Query March 94.2%; Score 1481.2; DB 10; Length 1721;
Best Local Similarity 95.5%; Pred. No. 0;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;
```

```

1
|||||
Db 161 AGAAGTTCAGTGGCCAGCTCTACTGAGAGAAATGCTGTAGATATACAGACGAGTACTC 240
Qy 241 TCCAGCCACAGCCCGGTTTCAAGCTCTCTCCACCACTCAGAGAACAGATGTCACTTGGCC 300
Db 241 TCCAGCCACAGCCCGGTTTCAAGCTCTCTCCACCACTCAGAGAACAGATGTCACTTGGCC 300
Qy 301 CCGGCACAGGAAACAGCTTCAAGTTCAAGTCCACCTGGGGAGACAGATGTCACTTGGCTC 360
Db 301 CCGGCACAGGAAACAGCTTCAAGTTCAAGTCCACCTGGGGAGACAGATGTCACTTGGCTC 360
Qy 361 CCGATCACAGGACAGGACCTTGGGCTTCCACCACTCCGCGACCCAGATGTCACTTGGCC 420
Db 361 CCGATCACAGGACAGGACCTTGGGCTTCCACCACTCCGCGACCCAGATGTCACTTGGCC 420
Qy 421 CCGGACAAACAGGACCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 480
Db 421 CCGGACAAACAGGACCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 480
Qy 481 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCC----- 513
Db 481 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 540
Qy 514 ----- 540
Db 541 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 600
Qy 541 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 600
Db 601 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 660
Qy 601 CCGGACACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 660
Db 661 CCGGACAAACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 720
Qy 661 CCGGACAAACAGGCGCGCCCGGGCTTCCACGCGCCCGCCCGACGATGTCACTTGGCC 720
Db 721 TCAAGCTCTGATCAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTTGGCGAGGCT 780
Qy 721 TCAAGCTCTGATCAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTTGGCGAGGCT 780
Db 781 ACCACAAACCCGACAGGACAGGACCTCATTCTCAATTCACAGCCACCTCTGATACT 840
Qy 781 ACCACAAACCCGACAGGACAGGACCTCATTCTCAATTCACAGCCACCTCTGATACT 840
Db 841 CCTACCAACCCCTGACAGCATGTGACACAAAGTGTGACAGTACCTCACATAGACAG 900
Qy 841 GATACCTCTCTCAGCTCTCTCTCCATCAGACGACTTCTCCAGTGTACTAGGGGTCT 900
Db 901 GATACCTCTCTCAGCTCTCTCTCCATCAGACGACTTCTCCAGTGTACTAGGGGTCT 960
Qy 901 TTTCTTTTCTGCTCTTTTCAATTTCAACTCAGTTTATTTCTCTCTGAAATGCC 960
Db 961 TTTCTTTTCTGCTCTTTTCAATTTCAACTCAGTTTATTTCTCTCTGAAATGCC 1020
Qy 961 AGCACCGACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTAT 1020
Db 1021 AGCACCGACTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTAT 1080
Qy 1021 AAAACAAGGGGGTTTTCTGGGGCTCTCCAAATTATTAAGTTCAAGGACAGAACTGTGGTA 1080
Db 1081 AAAACAAGGGGGTTTTCTGGGGCTCTCCAAATTATTAAGTTCAAGGACAGAACTGTGGTA 1140
Qy 1081 CAATTGACTCTGGCTTCCGAGAAAGTACATGAATTCAGAGTGTGAGAAACAATTC 1140
Db 1141 CAATTGACTCTGGCTTCCGAGAAAGTACATGAATTCAGAGTGTGAGAAACAATTC 1200
Qy 1141 AATCAGTATTAACGAGAGAGCTCTCGATATTAACCTGACGATCTCAGAGTCAAGCTG 1200
Db 1201 AATCAGTATTAACGAGAGAGCTCTCGATATTAACCTGACGATCTCAGAGTCAAGCTG 1260
Qy 1201 AGTCATGTGCATTTCTTTCTCTGCCAGTCTGGGGCTGGGGTGCAGAGCTGGGGCATC 1260

```

```

Db 1261 AGTGAATGCCATTTCTTTCTCTGCCAGTCTGGGGCTGGGGTGCAGAGCTGGGGCATC 1320
Qy 1261 GCGCTCTGTGCTGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1320
Db 1321 GCGCTCTGTGCTGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1380
Qy 1321 GCTGTCTGTCAAGTCCGCGGAAAGAACTACGCGGACGTGACATCTTTCCAGCCCGGAT 1380
Db 1381 GCTGTCTGTCAAGTCCGCGGAAAGAACTACGCGGACGTGACATCTTTCCAGCCCGGAT 1440
Qy 1381 ACCTACATCTTATGAGAGAGTACCCCACTTACCAACCCATGGGCGTATGTGCCCCCT 1440
Db 1441 ACCTACATCTTATGAGAGAGTACCCCACTTACCAACCCATGGGCGTATGTGCCCCCT 1500
Qy 1441 AGCAGTACCGATCTGATGAGCCCTTATGAGAGGTTTCTGAGATTAATGTGGAGAGCCTC 1500
Db 1501 AGCAGTACCGATCTGATGAGCCCTTATGAGAGGTTTCTGAGATTAATGTGGAGAGCCTC 1560
Qy 1501 TCTTACACAAACCCAGAGAGTGGAGGCACTTGTGCAACTTGTAGGGGACGTTGCCCTC 1560
Db 1561 TCTTACACAAACCCAGAGAGTGGAGGCACTTGTGCAACTTGTAGGGGACGTTGCCCTC 1620
Qy 1561 TGAGCTGAGTGG 1572
Db 1621 TGAGCTGAGTGG 1632

RESULT 3
US-09-964-824A-573
; Sequence 573, Application US/09964824A
; Patent No. US2002010251A1
; GENERAL INFORMATION:
; APPLICANT: Horigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-73
; CURRENT FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 573
; LENGTH: 1804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-573

Query Match 89.8%; Score 1412.2; DB 10; Length 1804;
Best Local Similarity 95.3%; Pred. No. 0;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

```


295 GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 354
QY ACCCGGCGAGCCAGATGTCACTCGAGCCCCGACACAGGCCCCGCGGCTCCAC 450
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 414
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 492
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 474
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 552
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 534
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 612
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 594
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 672
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 654
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 732
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 714
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 792
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 774
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 852
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 834
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 912
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 894
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 972
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 954
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1032
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1014
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1092
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1074
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1152
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1134
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1232
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1214
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1292
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1274
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1352
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1334
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1412
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1394
QY GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1492
Db GCGGCCCCAGCCAGGTGTCACTCGGCCCCGACACAGGCCCCGCGGCTCCAC 1474

1375 ATGAGCGAGTACCCCACTCAACACACAGGCGGTATGTCCCTTAGAGTACGAT 1434
QY GGTAGCCCTATGAGAGAGTTTGTGAGGTATGTGTGGCAGAGCCTCTTACACAAAC 1512
Db GGTAGCCCTATGAGAGAGTTTGTGAGGTATGTGTGGCAGAGCCTCTTACACAAAC 1494
QY CCAGCGGCGAGCCACTTGGCAGCACTTGTAG 1545
Db CCAGCGGCGAGCCACTTGGCAGCACTTGTAG 1527

RESULT 5
US-09-964-824A-105
Sequence 105, Application US/09964824A
Patent No. US2002010251A1
GENERAL INFORMATION:
APPLICANT: Horrihan, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-73
CURRENT FILING DATE: 2001-09-27
PRIOR FILING DATE: US/60/236,033
PRIOR FILING DATE: 2000-09-28
PRIOR FILING DATE: US/60/236,032
PRIOR FILING DATE: 2000-09-28
PRIOR FILING DATE: US/60/236,028
NUMBER OF SEQ ID NOS: 583
SOFTWARE: PatentIn version 3.0
SEQ ID NO 105
LENGTH: 4139
TYPE: DNA
ORGANISM: Homo sapiens
US-09-964-824A-105

Query Match 75.7%; Score 1190.2; DB 10; Length 4139;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 249 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGACAGAGTGTCACTGTGAGCCCGGACAC 308
Db 2545 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGACAGAGTGTCACTGTGAGCCCGGACAC 2604
QY 309 GAAACCAAGCTTCAAGTTCAGGTGTCACTGTGAGAGAGAGTGTCACTGTGAGCCCGGACAC 368
Db 2605 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2664
QY 369 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 428
Db 2665 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2724
QY 429 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 488
Db 2725 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2784
QY 489 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 548
Db 2785 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2844
QY 549 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 608
Db 2845 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2904
QY 609 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 668
Db 2905 CAGGCGCGGTTGAGGCTCTTCAACCACTCAGGAGAGAGAGTGTCACTGTGAGCCCGGACAC 2964
QY 669 TGCATGAGGCTTCACTGTGAGAGAGAGAGTGTCACTGTGAGCCCGGACAC 728
Db 2965 TGCATGAGGCTTCACTGTGAGAGAGAGAGTGTCACTGTGAGCCCGGACAC 3024
QY 729 CCGAGCGAGAGAGAGTGTCACTGTGAGAGAGAGTGTCACTGTGAGCCCGGACAC 788

Db 3025 CCCAGGACGACGAGGCACTCCATTTCTCAATTCACGACCACTCTGATCTTACCTTACAC 3084
Qy 789 CTTTGCACGACGATGACGACCAAGACTGATGACGATGACGATGACGATGACGATGACGATGAC 848
Db 3085 CTTTGCACGACGATGACGACCAAGACTGATGACGATGACGATGACGATGACGATGACGATGAC 3144
Qy 849 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 908
Db 3145 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3204
Qy 909 CTTGCTCTTTCACATTTCAAACTCCAGATTAATTCCTCTGAGAAATCCGACACCA 968
Db 3205 CTTGCTCTTTCACATTTCAAACTCCAGATTAATTCCTCTGAGAAATCCGACACCA 3264
Qy 969 CTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTTAAACAAG 1028
Db 3265 CTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTTAAACAAG 3324
Qy 1029 GGGTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGCGCAGATCTGTGTGTACATTTGAC 1088
Db 3325 GGGTTTTCTGGGCTCTCTCAATTAATTAAGTTCAGGCGCAGATCTGTGTGTACATTTGAC 3384
Qy 1089 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1148
Db 3385 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3444
Qy 1149 TAAACCGAAGAGCTCTGATATTAAGTTCAGATGACGATGACGATGACGATGACGATGAC 1208
Db 3445 TAAACCGAAGAGCTCTGATATTAAGTTCAGATGACGATGACGATGACGATGACGATGAC 3504
Qy 1209 GGCATTTCTTTCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1268
Db 3505 GGCATTTCTTTCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3564
Qy 1269 GGTGCTGTCTGTGTTCTGTTGCTGCTGACATGTTCTATCTCATTTGCTTGTCTGTG 1328
Db 3565 GGTGCTGTCTGTGTTCTGTTGCTGCTGACATGTTCTATCTCATTTGCTTGTCTGTG 3624
Qy 1329 TCAGTCCGCGAAAGAACTACGAGGACGTGACATCTTTCCAGCCGCGAATACCTACCA 1388
Db 3625 TCAGTCCGCGAAAGAACTACGAGGACGTGACATCTTTCCAGCCGCGAATACCTACCA 3684
Qy 1389 TCTATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1448
Db 3685 TCTATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3744
Qy 1449 CGATGCTAGCCCTTATGAGAAAGTTTCTGACAGTAAATGTTGACAGAGCTCTCTTACAC 1508
Db 3745 CGATGCTAGCCCTTATGAGAAAGTTTCTGACAGTAAATGTTGACAGAGCTCTCTTACAC 3804
Qy 1509 AAACCGAAGAGCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1547
Db 3805 AAACCGAAGAGCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3843

RESULT 6
US-09-964-824A-578
Sequence 578, Application US/09964824A
Patent No. US20020102531A1
GENERAL INFORMATION:
APPLICANT: Horigan, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening using Signatu
FILE REFERENCE: 689290-73
CURRENT APPLICATION NUMBER: US/09/964, 824A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US/60/236, 033
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 032
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236, 028
PRIOR FILING DATE: 2000-09-28

NUMBER OF SEQ ID NOS: 583
SOFTWARE: PatentIn version 3.0
SEQ ID NO 578
LENGTH: 4139
TYPE: DNA
ORGANISM: Homo sapiens
US-09-964-824A-578
Query Match 75.7%; Score 1190.2; DB 10; Length 4139;
Best Local Similarity 94.8%; Pred. No. 0; Mismatches 68; Indels 0; Gaps 0;
Matches 1331; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 249 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 308
Db 2545 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2604
Qy 309 GGAACCAAGCTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCA 368
Db 2605 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2664
Qy 369 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 428
Db 2665 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2724
Qy 429 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 488
Db 2725 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2784
Qy 489 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 548
Db 2785 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2844
Qy 549 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 608
Db 2845 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2904
Qy 609 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 668
Db 2905 CAGCCCGGTTTCAAGGCTCTCTCCACCACTCAGGACAGATGTCATCTGTGCCCCGAC 2964
Qy 669 TGCATCAGGCTCAGCTTCTACTGTGACACAGGACCTCTGACGAGCTACCAAC 728
Db 2965 TGCATCAGGCTCAGCTTCTACTGTGACACAGGACCTCTGACGAGCTACCAAC 3024
Qy 729 CCGAC 788
Db 3025 CCGAC 3084
Qy 789 CCGAC 848
Db 3085 CCGAC 3144
Qy 849 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 908
Db 3145 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3204
Qy 909 CTTGCTCTTTCACATTTCAAACTCCAGATTAATTCCTCTGAGAAATCCGACACCA 968
Db 3205 CTTGCTCTTTCACATTTCAAACTCCAGATTAATTCCTCTGAGAAATCCGACACCA 3264
Qy 969 CTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTTAAACAAG 1028
Db 3265 CTACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTCAGATTTTAAACAAG 3324
Qy 1029 GGGTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGCGCAGATCTGTGTGTACATTTGAC 1088
Db 3325 GGGTTTTCTGGGCTCTCTCAATTAATTAAGTTCAGGCGCAGATCTGTGTGTACATTTGAC 3384
Qy 1089 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 1148
Db 3385 TCTGACCTCTGACGATGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 3444
Qy 1149 TAAACCGAAGAGCTCTGATATTAAGTTCAGATGACGATGACGATGACGATGACGATGAC 1208

Db 3445 TAAACGGAAGCAGCTCTCATATTAACCTGACGATCTCAGACGTCAGCGTGAATGT 3504
 Qy 1209 GCCATTCTCTTCTGTGCGCAGTGTGGGGTGGGGGCGCAGGCTGGGGGCAATCGGGCTGT 1268
 Db 3505 GCCATTCTCTTCTGTGCGCAGTGTGGGGTGGGGGCGCAGGCTGGGGGCAATCGGGCTGT 3564
 Qy 1269 GGT 1328
 Db 3565 GGT 3624
 Qy 1329 TCAGTGCAGCGAAGAACTACGAGGAGCTGACATCTTCCAGCCGGGATACCTAACCA 1388
 Db 3625 TCAGTGCAGCGAAGAACTACGAGGAGCTGACATCTTCCAGCCGGGATACCTAACCA 3684
 Qy 1389 TCCTATGACGAGTACCCCACTACCAACCCATGGGCGCTATGTGCCCTAGCAGTAC 1448
 Db 3685 TCCTATGACGAGTACCCCACTACCAACCCATGGGCGCTATGTGCCCTAGCAGTAC 3744
 Qy 1449 CGATGTAGCCCTTATGAGAGTGTCTGAGGTATATGTGGGACAGACGCTCTTACAC 1508
 Db 3745 CGATGTAGCCCTTATGAGAGTGTCTGAGGTATATGTGGGACAGACGCTCTTACAC 3804
 Qy 1509 AAACCCAGAGTGGGAGCCACTTCTGCCAATTGTAGG 1547
 Db 3805 AAACCCAGAGTGGGAGCCACTTCTGCCAATTGTAGG 3843

RESULT 7
 US-09-864-864-334
 ; Sequence 334, Application US/09864864
 ; Patent No. US20020102679A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Dillon, Devin C.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Fling, Steve P.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Carter, Derrick
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.523
 ; CURRENT APPLICATION NUMBER: US/09/864,864
 ; NUMBER OF SEQ ID NOS: 341
 ; SOFTWARE: Corixa Invention Disclosure Database
 ; SEQ ID NO 334
 ; LENGTH: 4139
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-864-864-334

Query Match 75.7%; Score 1190.2; DB 10; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 0;
 Matches 1211; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 249 CAGCCCGGTTGAGGCTCTTCCACACTCAGGAGAGATGTCACTTGGCCCGGAC 308
 Db 2545 CAGGCGGCGCGCGGCTCCACCGCCCCCAGCCACGAGTGTCACTTGGCCCGGAC 2604
 Qy 309 GGAACGAGTTGAGTGTGAGTGTGACCTGGGAGAGATGTCACTTGGCCCGGAC 368
 Db 2605 CAGGCGGCGCGCGGCTCCACCGCCCCCAGCCACGAGTGTCACTTGGCCCGGAC 2664
 Qy 369 CAGGCGGCGCGGCTCCACCGCCCCCAGCCACGAGTGTCACTTGGCCCGGAC 428
 Db 2665 CAGGCGGCGCGGCTCCACCGCCCCCAGCCACGAGTGTCACTTGGCCCGGAC 2724

Qy 429 CAAGCAGCCCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 488
 Db 2725 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 2784
 Qy 489 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 548
 Db 2785 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 2844
 Qy 549 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 608
 Db 2845 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 2904
 Qy 609 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 668
 Db 2905 CAGGCGGCGCGGCTCCACCGCCCCCGGACCAAGTGTCACTTGGCCCGGAC 2964
 Qy 669 TGCAATCAGGCTCAGCTTCTACTCTGTGTGACCAAGGACCTCTGCCAGGAGCTAAC 728
 Db 2965 TGCAATCAGGCTCAGCTTCTACTCTGTGTGACCAAGGACCTCTGCCAGGAGCTAAC 3024
 Qy 729 CCGAGCCAGCAAGAGCACTCCACCGAGATTCGACGACCACTGTATCTCTTAC 788
 Db 3025 CCGAGCCAGCAAGAGCACTCCACCGAGATTCGACGACCACTGTATCTCTTAC 3084
 Qy 789 CCGTGCAGGCAATGAGCAAGAGATGATGAGGAGCACTACCATGAGAGCTTCC 848
 Db 3085 CCGTGCAGGCAATGAGCAAGAGATGATGAGGAGCACTACCATGAGAGCTTCC 3144
 Qy 849 TCTCACTCTCTCCATCAGAGCACTTCTCCAGTGTGTATCTGGGGTCTCTTTT 908
 Db 3145 TCTCACTCTCTCCATCAGAGCACTTCTCCAGTGTGTATCTGGGGTCTCTTTT 3204
 Qy 909 CCTGTCTTTTCACTTCAAACTTCAGTTTATTTCTCTGGAAGATCCAGAGCA 968
 Db 3205 CCTGTCTTTTCACTTCAAACTTCAGTTTATTTCTCTGGAAGATCCAGAGCA 3264
 Qy 969 CTACTCAAGAGCTGAGAGAGACATTTGAAATGTTTTGAGATTTATAAAGG 1028
 Db 3265 CTACTCAAGAGCTGAGAGAGACATTTGAAATGTTTTGAGATTTATAAAGG 3324
 Qy 1029 GGGTTTTCTGGGCTCTCCAAATTAATTAAGTCAAGGAGATCTGTGGTGTCAATGAC 1088
 Db 3325 GGGTTTTCTGGGCTCTCCAAATTAATTAAGTCAAGGAGATCTGTGGTGTCAATGAC 3384
 Qy 1089 TCTGGCTTCCGAAAGATGATCATGTTCACAGCTGAGACACAGTTCAATGAT 1148
 Db 3385 TCTGGCTTCCGAAAGATGATCATGTTCACAGCTGAGACACAGTTCAATGAT 3444
 Qy 1149 TAAACGAGAGAGCTCTCGATTAATTAATGATCTGAGAGTGTCAAGTGTGATGT 1208
 Db 3445 TAAACGAGAGAGCTCTCGATTAATTAATGATCTGAGAGTGTCAAGTGTGATGT 3504
 Qy 1209 GCCATTCTCTTCTGTGCGCAGTGTGGGGTGGGGGCGCAGGCTGGGGGCAATCGGGCTGT 1268
 Db 3505 GCCATTCTCTTCTGTGCGCAGTGTGGGGTGGGGGCGCAGGCTGGGGGCAATCGGGCTGT 3564
 Qy 1269 GGT 1328
 Db 3565 GGT 3624
 Qy 1329 TCAGTGCAGCGAAGAACTACGAGGAGCTGACATCTTCCAGCCGGGATACCTAACCA 1388
 Db 3625 TCAGTGCAGCGAAGAACTACGAGGAGCTGACATCTTCCAGCCGGGATACCTAACCA 3684
 Qy 1389 TCCTATGACGAGTACCCCACTACCAACCCATGGGCGCTATGTGCCCTAGCAGTAC 1448
 Db 3685 TCCTATGACGAGTACCCCACTACCAACCCATGGGCGCTATGTGCCCTAGCAGTAC 3744
 Qy 1449 CGATGTAGCCCTTATGAGAGTGTCTGAGGTATATGTGGGACAGACGCTCTTACAC 1508
 Db 3745 CGATGTAGCCCTTATGAGAGTGTCTGAGGTATATGTGGGACAGACGCTCTTACAC 3804
 Qy 1509 AAACCCAGAGTGGGAGCCACTTCTGCCAATTGTAGG 1547


```

? LOCATION: (55)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (1045)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (1355)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-19

```

| | | | | |
|---------------------------|--------|---------------------|-----------|--------------|
| Query Match | 62.2%; | Score 978; | DB 10; | Length 1355; |
| Best Local Similarity | 98.9%; | Pred. No. 5.8e-287; | | |
| Matches 984; Conservative | 0; | Mismatches 11; | Indels 0; | Gaps 0; |

| | | | |
|----|------|---|------|
| Qy | 578 | CGAGCCACAGAGTGCATCTCGGAGCCCGGAGCAACAAGGCGGGCTTTGGGCTCTCAACGCCCTTC | 637 |
| Db | 1 | CAGCCCATGGTGTACACTCTGGAGCCCGGAGCAACAAGCCCGCTTTGGGCTCTCAACGCCCTTC | 60 |
| Qy | 638 | CAGTGCACATGTTCACACTCTGGGCTCAGAGCTCTGCATCAGAGCTCAGCTTCTACTCTGTGGC | 697 |
| Db | 61 | CAGTGCACAAATGTACACTCTGGGCTCAGAGGCTCTGCATCAGAGCTCAGCTTCTACTCTGTATC | 120 |
| Qy | 698 | ACAAAGGACACTCTGCGACAGGAGCTACCAAAACCCAGCGACAAAGACACTTCACCCAGCA | 757 |
| Db | 121 | ACAAAGGACACTCTGCGACAGGAGCTACCAAAACCCAGCGACAAAGACACTTCATTTCTCAA | 180 |
| Qy | 758 | TTTCGAGCGACCACTGTATATCTCTTACCAACCTTCGCGACGATATAGCACCAAGACTGAAG | 817 |
| Db | 181 | TTTCGAGCGACCACTGTATATCTCTTACCAACCTTCGCGACGATATAGCACCAAGACTGAAG | 240 |
| Qy | 818 | CCGATAGCATCTACACCTTACACAGGTAACCTCTCTCTACCTCTCTCATATCAGACGACTTTC | 877 |
| Db | 241 | CCGATAGCATCTACACCTTACACAGGTAACCTCTCTCTACCTCTCTCATATCAGACGACTTTC | 300 |
| Qy | 878 | CCCAATGTCTACTCGGAGCTCTCTTTCTTTTCTGTCTTTTCCATTTTCAAACTCCAGT | 937 |
| Db | 301 | CCCAATGTCTACTCGGAGCTCTCTTTCTTTTCTGTCTTTTCCATTTTCAAACTCCAGT | 360 |
| Qy | 938 | TTATATTTCTCTCTGAAAGATCCGACGACGACATACCTAACAGAGCTGCGAGAGACACTTT | 997 |
| Db | 361 | TTATATTTCTCTCTGAAAGATCCGACGACGACATACCTAACAGAGCTGCGAGAGACACTTT | 420 |
| Qy | 998 | CTGAATCTTTTTCGAGATTTATTAACAAGAGGGTTTCTGGGCTCTCTCAATTTTAAGT | 1057 |
| Db | 421 | CTGAATCTTTTTCGAGATTTATTAACAAGAGGGTTTCTGGGCTCTCTCAATTTTAAGT | 480 |
| Qy | 1058 | TCAGGCCAGATCTGTGTGTGATCAATTTGACTCTGAGCTTCTCGAAGAGTACATCAATG | 1117 |
| Db | 481 | TCAGGCCAGATCTGTGTGTGATCAATTTGACTCTGAGCTTCTCGAAGAGTACATCAATG | 540 |
| Qy | 1118 | TCGACGACGTGTGAGACACAGTTCATCAGTATTAACAAGAAAGAGCTCTCGATTAATAC | 1177 |
| Db | 541 | TCGACGACGTGTGAGACACAGTTCATCAGTATTAACAAGAAAGAGCTCTCTCGATTAATAC | 600 |
| Qy | 1178 | TGACGATCTCGACAGTCAACCGTAGATCAATGTGATGACATTTCTTCTCTCGACAGTCTGGG | 1237 |
| Db | 601 | TGACGATCTCGACAGTCAACCGTAGATCAATGTGATGACATTTCTTCTCTCGACAGTCTGGG | 660 |
| Qy | 1238 | CTGGGAGTCCGAGCTGGGACATCGCGCTGCTGATCTGTCTGTGTTCTGTGGTGCCTGG | 1297 |
| Db | 661 | CTGGGAGTCCGAGCTGGGACATCGCGCTGCTGATCTGTCTGTGTTCTGTGGTGCCTGG | 720 |
| Qy | 1298 | CCATGTCTATCTCATTTGCTCTGAGCTCTCTGTACATGTCGCGCCGAAAGAACTACGGGAGC | 1357 |
| Db | 721 | CCATGTCTATCTCATTTGCTCTGAGCTCTCTGTACATGTCGCGCCGAAAGAACTACGGGAGC | 780 |
| Qy | 1358 | TGAGCATCTTTCCAGGCGGGAGATACCTACATCTCTATAGAGATGATCCGACCTACACACA | 1417 |
| Db | 781 | TGAGCATCTTTCCAGGCGGGAGATACCTACATCTCTATAGAGATGATCCGACCTACACACA | 840 |
| Qy | 1418 | CCCATGGGCGTATGTGCCCCCTTACGACATACGATCTGATGAGCCCTTATGAGAAAGTTTCTG | 1477 |
| Db | 841 | CCCATGGGCGTATGTGCCCCCTTACGACATACGATCTGATGAGCCCTTATGAGAAAGTTTCTG | 900 |

| | | | |
|----|------|--|------|
| Oy | 1478 | CAGGTAATGGTGGCAGACAGCCTCTCTTCAACAAACGAGAGTGGACAGCACTTTCGCA | 1537 |
| Db | 901 | CAGGTAATGGTGGCAGACAGCCTCTCTTCAACAAACGAGAGTGGACAGCACTTTCGCA | 960 |
| Oy | 1538 | ACTTGTGAGGGGCACGTGCTCCCTCTGAGCTGAGATGG | 1572 |
| Db | 961 | ACTGTATAGGGGCACGTGCTCCCTCTGAGCTGAGATGG | 995 |

RESULT 10
US-09-864-864-194
; Sequence 194, Application US/09864864
; Patent No. US20020102679A1
GENERAL INFORMATION

APPLICANT: Xu, Jiangchun
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Dillon, David C.
APPLICANT: Secrist, Heather
APPLICANT: Lodes, Michael J.
APPLICANT: Alagste, Paul A.
APPLICANT: Pfling, Steve P.
APPLICANT: Mannheim, Jane R.
APPLICANT: Benson, Darin R.
APPLICANT: Carter, Derrick
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF OVARIAN CANCER
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.523
CURRENT APPLICATION NUMBER: US/09/864,864
CURRENT FILING DATE: 2001-05-23
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 194
LENGTH: 497
TYPE: DNA
ORGANISM: Homo sapiens
US-09-864-864-194

| | | | | |
|---------------------------|--------|-------------------|-----------|-------------|
| Query Match | 31.5%; | Score 494.4; | DB 10; | length 497; |
| Best Local Similarity | 99.8%; | Pred. No. 4e-140; | | |
| Matches 495; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps 0; |

| | | | |
|----|------|--|------|
| QY | 771 | CTGTGATATCTTCACACACCTTGGACGCAATMGACCAABAAGTGAAGCAATGACACTAC | 830 |
| Db | 2 | CTCTATATCTCTACACACCTTGGACGCAATMGACCAABAAGTGAAGCAATGACACTAC | 61 |
| QY | 831 | CCATGACAGGTAACCTCTCTACCTCTCCATACACAGACCTTCCCGAGTTGTAC | 890 |
| Db | 62 | CCATGACAGGTAACCTCTCTACCTCTCCATACACAGACCTTCCCGAGTTGTAC | 121 |
| QY | 891 | TGGGGTCTCTTCTCTTCTCTGTCTTTTGCATTTCAAACTCCAGTTTAATCTCTCT | 950 |
| Db | 122 | TGGGGTCTCTTCTCTTCTCTGTCTTTTGCATTTCAAACTCCAGTTTAATCTCTCT | 181 |
| QY | 951 | GGAAGATCCGACGACCGACTCTACCAAGAGCTGACAGAGACATTTCTGAATGTTTT | 1010 |
| Db | 182 | GGAAGATCCGACGACCGACTCTACCAAGAGCTGACAGAGACATTTCTGAAGTTTTT | 241 |
| QY | 1011 | GCAGATTATTAACAAGGGGGGTTTTTGGGCGCTCCCAATATTAATTCAAGCCAGGATC | 1070 |
| Db | 242 | GCAGATTATTAACAAGGGGGGTTTTTGGGCGCTCCCAATATTAATTCAAGCCAGGATC | 301 |
| QY | 1071 | TGTGTGTATCAATTGACTGTGCGCTTCGAGAGAGTTACATCAATGTTCACGACGTGA | 1130 |
| Db | 302 | TGTGTGTGTAAATTAATCACTGTGGGCTTCGAGAGAGTTACATCAATGTTCACGACGTGA | 361 |
| QY | 1131 | GACACAGTTCAATCAATATATAAAGGAAGACCTCTGATATATACTTGACGATCTCAG | 1190 |
| Db | 362 | GACACAGTTCAATCAATATATAAAGGAAGACCTCTGATATATACTTGACGATCTCAG | 421 |
| QY | 1191 | CGTCAAGCTGATCATGTGCCATTTCTTTCTCCCAATCTGGGGCTGGGGGTCCAGG | 1250 |

Thu May, 8 18:53:09 2003

us-09-658-621b-1.inp

Page 1

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 15:04:32 ; Search time 3749 Seconds
(without alignments)
10542.542 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572
Sequence: 1 gaattccctgcgtcgtgaa.....tcgacctcgcagctgagtcgg 1572

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2479104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending Patents: NA, Main: *

1: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US0986_COMB.seq.*

| Result No. | Score | Query Match Length | ID | Description |
|------------|--------|--------------------|-----------------------------|---------------------|
| 1 | 1572 | 100.0 | 1572 25 US-09-658-621-1 | Sequence 1, Appli |
| 2 | 1481.8 | 94.3 | 2026 41 US-10-198-846-14589 | Sequence 12589, A |
| 3 | 1481.2 | 94.2 | 1721 1 PCT-US02-07826-211 | Sequence 211, App |
| 4 | 1481.2 | 94.2 | 1721 1 PCT-US02-18947-775 | Sequence 775, App |
| 5 | 1481.2 | 94.2 | 1721 33 US-09-864-864-280 | Sequence 280, App |
| 6 | 1481.2 | 94.2 | 1721 36 US-09-967-768A-224 | Sequence 224, App |
| 7 | 1481.2 | 94.2 | 1721 38 US-10-007-926A-58 | Sequence 58, Appli |
| 8 | 1481.2 | 94.2 | 1721 39 US-10-029-517-3 | Sequence 3, Appli |
| 9 | 1481.2 | 94.2 | 1721 39 US-10-097-340-211 | Sequence 211, App |
| 10 | 1481.2 | 94.2 | 1721 41 US-10-172-118-775 | Sequence 155, App |
| 11 | 1481.2 | 94.2 | 1721 41 US-10-172-118-775 | Sequence 775, App |
| 12 | 1481.2 | 94.2 | 2678 42 US-10-252-157-103 | Sequence 103, App |
| 13 | 1481.2 | 94.2 | 2678 42 US-10-252-157-103 | Sequence 103, App |
| 14 | 1481.2 | 94.2 | 2574 28 US-09-716-473-2290 | Sequence 2290, App |
| 15 | 1454.4 | 92.5 | 2574 29 US-09-726-172-2117 | Sequence 2117, App |
| 16 | 1454.4 | 92.5 | 2574 29 US-09-726-172-2117 | Sequence 2117, App |
| 17 | 1412.2 | 89.8 | 1804 36 US-10-029-517-17 | Sequence 573, Appli |
| 18 | 1412.2 | 89.8 | 1804 36 US-10-029-517-17 | Sequence 339, Appli |
| 19 | 1342 | 85.4 | 1428 1 PCT-US02-08456-339 | Sequence 17, Appli |
| 20 | 1309.8 | 83.3 | 1452 8 US-08-479-537-1 | Sequence 2, Appli |
| 21 | 1289.2 | 82.0 | 1443 8 US-08-479-537-2 | Sequence 2, Appli |

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

```
22 1257 80.0 1527 17 US-09-366-670-19 Sequence 19, Appl
23 1257 80.0 1527 39 US-10-057-136-19 Sequence 19, Appl
24 1213.4 77.2 1737 32 US-09-830-513-11 Sequence 11, Appl
25 1210.2 77.0 1371 32 US-09-830-513-5 Sequence 5, Appl
26 1190.2 75.7 4139 1 PCT-US02-19669-310 Sequence 310, Appl
27 1190.2 75.7 4139 33 US-09-864-864-334 Sequence 2121, Ap
28 1190.2 75.7 4139 33 US-09-880-107-2121 Sequence 105, App
29 1190.2 75.7 4139 36 US-09-964-824A-105 Sequence 157, App
30 1190.2 75.7 4139 41 US-10-171-311-157 Sequence 310, App
31 1190.2 75.7 4139 41 US-10-171-293-310 Sequence 5270, Ap
32 1188.6 75.6 4139 65 US-09-213-360-5270 Sequence 2741, Ap
33 1188.6 75.5 4168 16 US-09-205-070-2741 Sequence 2741, Ap
34 1186.8 75.5 4168 17 US-09-340-623-2741 Sequence 2741, Ap
35 1186.8 75.5 4168 37 US-09-898-888-2741 Sequence 2741, Ap
36 1186.8 75.5 4168 33 US-09-898-888A-2741 Sequence 41, Appl
37 1043.6 66.4 2297 32 US-09-856-988A-41 Sequence 19, Appl
38 978 62.2 1355 1 PCT-US00-05882-19 Sequence 19, Appl
39 978 62.2 1355 34 US-09-925-301-19 Sequence 7468, Ap
40 946.2 60.2 1261 28 US-09-705-256A-7468 Sequence 10940, A
41 946.2 60.2 1261 60 US-60-164-285-7468 Sequence 1, Appl1
42 919.4 58.5 3361 71 US-60-278-258-10840 Sequence 19, Appl1
43 750 47.7 8186 38 US-10-029-517-19 Sequence 19, Appl1
44 750 47.7 8186 38 US-10-029-517-19 Sequence 19, Appl1
45 750 47.7 8186 38 US-10-029-517-19 Sequence 19, Appl1
```

ALIGNMENTS

```
RESULT 1
US-09-658-621-1
; Sequence 1, Application US/09658621
; GENERAL INFORMATION:
; APPLICANT: Taylor-Papadimitriou, Joyce
; APPLICANT: Henkamp, Lukas Carl
; APPLICANT: Offitger, Rieck
; APPLICANT: Melief, Cornelis Johanna Maria
; APPLICANT: Acres, Bruce
; APPLICANT: Thomas, Mireille
; TITLE OF INVENTION: MUC-1 derived peptides
; FILE REFERENCE: 029395-017
; CURRENT APPLICATION NUMBER: US/09/658, 621
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/187,215
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: GB 9921242.5
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: EP 99 40 2237.4
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)..(1542)
US-09-658-621-1
```

Query Match 100.0%; Score 1572; DB 25; Length 1572;

Best Local Similarity 100.0%; Pred. No. 8.9e-296; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCCTGCTCTTGAATCTGTCGCCCCCTCCCAACCCATTTCACCAACCATG 60
DB 1 GAATTCCTGCTCTTGAATCTGTCGCCCCCTCCCAACCCATTTCACCAACCATG 60
QY 61 AACACGGGACACCAAGTCTCTTCTCTGCTGCTCTCTCAAGTGTGTTT 120
DB 61 AACACGGGACACCAAGTCTCTTCTCTGCTGCTCTCTCAAGTGTGTTT 120
QY 121 AACAGTTCGTGATGCAAGCTACCCAGGTGGAAGAAAGAGACTTCGGCTACCCAG 180
DB 121 AACAGTTCGTGATGCAAGCTACCCAGGTGGAAGAAAGAGACTTCGGCTACCCAG 180
```

```
|||||
DB 121 ACAGGTTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCTACCCAG 180
QY 181 AGAAGTTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCTACCCAG 240
DB 181 AGAAGTTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCTACCCAG 240
QY 241 TCCAGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 300
DB 241 TCCAGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 300
QY 301 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 360
DB 301 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 360
QY 361 CCAAGTCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 420
DB 361 CCAAGTCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 420
QY 421 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 480
DB 421 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 480
QY 481 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 540
DB 481 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 540
QY 541 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 600
DB 541 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 600
QY 601 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 660
DB 601 CCGGCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 660
QY 720 TCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 720
DB 720 TCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 720
QY 780 ACCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 780
DB 780 ACCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 780
QY 840 CCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 840
DB 840 CCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 840
QY 900 GCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 900
DB 900 GCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 900
QY 960 TCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 960
DB 960 TCAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 960
QY 1020 AGAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1020
DB 1020 AGAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1020
QY 1080 AAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1080
DB 1080 AAAGGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1080
QY 1140 CAATGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1140
DB 1140 CAATGCTCTGATGCAAGCTCTACCCAGGTGGAAGAAAGAGACTTCGGCT 1140
QY 1200 AATCAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1200 AATCAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1260 AGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1260 AGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
```



```
Db 1201 AGTCATGTGCATTTCTTTCTCTGACCAGTCTGGGGCTGGGGGTGCCAGGCTGGGGCATC 1260
Qy 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Qy 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Qy 1381 ACCATCCATCCATGAGAGAGATACCCGACCTGACACACCCATGAGGAGCTATGAGCCCTC 1440
Db 1381 ACCATCCATCCATGAGAGAGATACCCGACCTGACACACCCATGAGGAGCTATGAGCCCTC 1440
Qy 1441 AGCATGACCGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 1441 AGCATGACCGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Qy 1501 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1501 TCTTACCAAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1561 TGAGCTGAGTGG 1572
Db 1561 TGAGCTGAGTGG 1572

RESULT 2
US-10-198-846-12589
; Sequence 12589, Application US/10198846
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OR INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OR INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OR INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCES: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12589
; LENGTH: 2026
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1, 2026
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12589

Query Match 94.3%; Score 1481.8; DB 41; Length 2026;
Best Local Similarity 95.6%; Pred. No. 4,3e-280;
Matches 1559; Conservative 0; Mismatches 12; Indels 60; Gaps 1;
```

```
Db 190 GAAGTTACGTGCCAGCTCTACTGAGAAATGCTGTGATGATGATGATGATGATGATGATGAT 249
Qy 242 CCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
Db 250 CCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Qy 302 CCGACACGGAACAGCTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAG 361
Db 310 CCGACACGGAACAGCTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAG 369
Qy 362 CAGTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Db 370 CAGTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
Qy 422 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Db 430 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
Qy 482 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 513
Db 490 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 549
Qy 514 -----CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
Db 550 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
Qy 542 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
Db 610 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
Qy 602 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
Db 670 CCGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
Qy 662 CAGGCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721
Db 730 CAGGCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
Qy 722 CCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
Db 790 CCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 849
Qy 782 CTACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
Db 850 CTACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 909
Qy 842 TACCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Db 910 TACCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 969
Qy 902 TCTTTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
Db 970 TCTTTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
Qy 962 GCACGAGACTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
Db 1030 GCACGAGACTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
Qy 1022 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 1090 AACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
Qy 1082 AATTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 1150 AATTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
Qy 1142 ATCAGATTAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db 1210 ATCAGATTAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269
Qy 1202 GTGATGTGCAATTTCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
Db 1270 GTGATGTGCAATTTCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1329
```


Qy 1201 AGTCATGTCATTTCTTCTCTGCGCAGTCGAGGCTGAGGTCAGAGCTGAGGATC 1260
|||
Db 1261 AGTCATGTCATTTCTTCTCTGCGCAGTCGAGGCTGAGGTCAGAGCTGAGGATC 1320
Qy 1261 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
|||
Db 1321 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
|||
Qy 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
|||
Db 1381 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
|||
Qy 1381 ACCTCATCCTCATGAGAGAGTACCCCATCTACCAACCCATGAGGCTATGAGCCCT 1440
|||
Db 1441 ACCTCATCCTCATGAGAGAGTACCCCATCTACCAACCCATGAGGCTATGAGCCCT 1500
|||
Qy 1441 AGCATGATCCGATCTGATCCCTCTATGAGAGGTTTCTGAGGATATGCTGAGCAGCTC 1500
|||
Db 1501 AGCATGATCCGATCTGATCCCTCTATGAGAGGTTTCTGAGGATATGCTGAGCAGCTC 1560
|||
Qy 1501 TCTTACACAAACCCAGAGATGAGCAGCCTTCTGCACTTGTATGAGGAGCAGTGCCTC 1560
|||
Db 1561 TCTTACACAAACCCAGAGATGAGCAGCCTTCTGCACTTGTATGAGGAGCAGTGCCTC 1620
|||
Qy 1561 TAGAGCTGAGTGG 1572
|||
Db 1621 TAGAGCTGAGTGG 1632
|||

RESULT 4
PCT-US02-18947-775
; Sequence 775, Application PC/TUS0218947
; GENERAL INFORMATION:
; APPLICANT: Rosetta Inpharmatics
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-228
; CURRENT APPLICATION NUMBER: PCT/US02/18947
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ. ID NOS: 2699
; SEQ. ID NO: 775
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002456
; DATABASE ENTRY DATE: 2001-06-18
; PCT-US02-18947-775

Query Match 94.2%; Score 1481.2; DB 1; Length 1721;
Best Local Similarity 95.5%; Pred. No. 5.5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

Qy 1 GAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
|||
Db 1 GAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
|||
Qy 61 ACACCGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
|||
Db 61 ACACCGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
|||
Qy 121 ACAGTTCTGCTCATGCAAGCTCTACACCGAGTGAAGAAAGAGAGCTTGGCTACCCAG 180
|||
Db 121 ACAGTTCTGCTCATGCAAGCTCTACACCGAGTGAAGAAAGAGAGCTTGGCTACCCAG 180
|||
Qy 181 AGAAGTTCAAGTCCCAAGCTCTACAGAGAAAGATGCTGAGATGATGACACAGAGCTACTC 240
|||
Db 181 AGAAGTTCAAGTCCCAAGCTCTACAGAGAAAGATGCTGAGATGATGACACAGAGCTACTC 240
|||
Qy 241 TCCAGCCACAGCCCGGCTTCAAGCTCTCTCCACCACTCAGAGGACAGATGTCATCTGAGCC 300
|||
Db 241 TCCAGCCACAGCCCGGCTTCAAGCTCTCTCCACCACTCAGAGGACAGATGTCATCTGAGCC 300
|||

Qy 301 CCGAGCAGAGAACCAAGCTTCAGATTCAAGCTGCAACCTGAGGAGACAGAGATGCACTCGGTC 360
|||
Db 301 CCGAGCAGAGAACCAAGCTTCAGATTCAAGCTGCAACCTGAGGAGACAGAGATGCACTCGGTC 360
|||
Qy 361 CCAATGACAGAGGCAAGCTGCTGAGCTTCCACCAACCCCGCAGGCCACAGATGCACTGAGCC 420
|||
Db 361 CCAATGACAGAGGCAAGCTGCTGAGCTTCCACCAACCCCGCAGGCCACAGATGCACTGAGCC 420
|||
Qy 421 CCGAGCAACAAAGCAGCCCGGAGCTTCCACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 480
|||
Db 421 CCGAGCAACAAAGCAGCCCGGAGCTTCCACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 480
|||
Qy 481 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 513
|||
Db 481 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 540
|||
Qy 514 -----CCGCGGCTTCAAGGATGCACTGAGCC 540
|||
Db 541 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 600
|||
Qy 541 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 600
|||
Db 601 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 660
|||
Qy 601 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 660
|||
Db 661 CCGAGCAACAGGCGGCGGCGGCGGCTGCAACCGGCTGCAACCGGCTGCAACCGGCTGCACTG 720
|||
Qy 661 TAGGCTCTGATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
|||
Db 721 TAGGCTCTGATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
|||
Qy 721 ACCACAAACCCAGCAGAGAGAGCACTCCACAGATTCAGAGCCACCACTGATACT 780
|||
Db 781 ACCACAAACCCAGCAGAGAGAGCACTCCACAGATTCAGAGCCACCACTGATACT 840
|||
Qy 781 CCTACACCTTGGAGCCATGAGACCAAGACCTGATGAGTACCTACCAATGAGACG 840
|||
Db 841 CCTACACCTTGGAGCCATGAGACCAAGACCTGATGAGTACCTACCAATGAGACG 900
|||
Qy 841 GTACCTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
|||
Db 901 GTACCTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
|||
Qy 901 TCTTTTCCGCTTTCATCAATTCATCAACCTGATTTATTCCTCTGAGAGATGCC 960
|||
Db 961 TCTTTTCCGCTTTCATCAATTCATCAACCTGATTTATTCCTCTGAGAGATGCC 1020
|||
Qy 961 AGACCGACTACTACCAAGAGCTGCAAGAGACATTTCTGGAATGTTTGCAGATTTAT 1020
|||
Db 1021 AGACCGACTACTACCAAGAGCTGCAAGAGACATTTCTGGAATGTTTGCAGATTTAT 1080
|||
Qy 1021 AAACAGGAGGTTTCTGAGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1080
|||
Db 1081 AAACAGGAGGTTTCTGAGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1140
|||
Qy 1081 CAATGACTGAGCTTCCGAGAGAGTACATCAATGTCACAGAGTGAAGACAGATTC 1140
|||
Db 1141 CAATGACTGAGCTTCCGAGAGAGTACATCAATGTCACAGAGTGAAGACAGATTC 1200
|||
Qy 1141 AATGATATTAACAGAGAGAGCTCTGATATTAACCTGAGATCTCAGAGCTGAGCTG 1200
|||
Db 1201 AATGATATTAACAGAGAGAGCTCTGATATTAACCTGAGATCTCAGAGCTGAGCTG 1260
|||
Qy 1201 AGTCATGTCATTTCTTCTCTGCGCAGTCGAGGCTGAGGTCAGAGCTGAGGATC 1260
|||
Db 1261 AGTCATGTCATTTCTTCTCTGCGCAGTCGAGGCTGAGGTCAGAGCTGAGGATC 1320
|||
Qy 1261 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
|||
Db 1321 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
|||

| | | | |
|----|------|--|------|
| Qy | 1321 | GCCTCTCTCAGAGCCGCCGAAAGACTACCGGAGCTGACACATCTTTTTCAGAGCCCGGAAAT | 1380 |
| Db | 1381 | GCCTCTCTCAGAGCCGCCGAAAGACTACCGGAGCTGACACATCTTTTTCAGAGCCCGGAAAT | 1440 |
| Qy | 1381 | ACCTACCATCTCTTATGAGGAGATACCCACCTGACACACCAATGGGGGCTATGATGCCCTCT | 1440 |
| Db | 1441 | ACCTACCATCTCTTATGAGGAGATACCCACCTGACACACCAATGGGGGCTATGATGCCCTCT | 1500 |
| Qy | 1441 | AGAGATACCAATGGTACGCCCTTATGAAABATTTTCTGATGATTAATCTGAGGAGAGACATCTC | 1500 |
| Db | 1501 | AGAGATACCAATGGTACGCCCTTATGAAABATTTTCTGATGATTAATCTGAGGAGAGACATCTC | 1560 |
| Qy | 1501 | TCTTTCACAAACCCACACATGACAGCACTTCTGCAACTTTTATAGGGGACAGTGGCCCTC | 1560 |
| Db | 1561 | TCTTTCACAAACCCACACATGACAGCACTTCTGCAACTTTTATAGGGGACAGTGGCCCTC | 1620 |
| Qy | 1561 | TGAGCTGAGTGG 1572 | |
| Db | 1621 | TGAGCTGAGTGG 1632 | |

RESULT 5
US-09-864-864-280
; Sequence 280, Application US/09864864
; GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Dillon, David C.
 APPLICANT: Secrist, Heather
 APPLICANT: Lodes, Michael J.
 APPLICANT: Algate, Paul A.
 APPLICANT: Fling, Steve P.
 APPLICANT: Mannion, Jane P.
 APPLICANT: Benson, Darin R.
 APPLICANT: Career: Derrick
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
 FILE REFERENCE: 210121.523
 CURRENT APPLICATION NUMBER: US/09/864, 864
 NUMBER OF SEO ID NOS: 341
 NUMBER OF SEO ID NOS: 2001-05-23
 SOFTWARE: Corixa Invention Disclosure Database
 SEO ID NO: 280
 LENGTH: 1721
 TYPE: DNA
 ORGANISM: Homo sapiens
 JS-09-864-864-280

| Query Match | 94.2%; | Score 1481.2; | DB 33; | Length 1721; |
|---------------------------|--------|---------------|--------|--------------|
| Best Total Classification | 94.2% | 1481.2 | 33 | 1721 |

Best local similarity: 95.5%; Pred. NO. 5,5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

| | | | |
|----|-----|--|-----|
| Qy | 1 | GAATTCCTGGCTGCTGAAATCTGTTTGGCCCCCTCCACCAATTTACACACACANG | 60 |
| Db | 1 | GAATTCCTGGCTGCTGAAATCTGTTTGGCCCCCTCCACCAATTTACACACACANG | 60 |
| Qy | 61 | AACACGGGCAACCAAGTCTCTTTCTTCTGGCTGCTCTTCACAGGCTTACAGTGT | 120 |
| Db | 61 | AACACGGGCAACCAAGTCTCTTTCTTCTGGCTGCTCTTCACAGGCTTACAGTGT | 120 |
| Qy | 121 | AAGAGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGAGCTGTTGGCTACAG | 180 |
| Db | 121 | AAGAGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGAGCTCTGGGCTACCAAG | 180 |
| Qy | 181 | AAGAAATTCAGTCCACAGCTCTCTAGAGAAAGAAATCTGTGATATGACCAAGACGTTATC | 240 |
| Db | 181 | AAGAAATTCAGTCCACAGCTCTCTAGAGAAAGAAATCTGTGATATGACCAAGACGTTATC | 240 |
| Qy | 241 | TTCAGCCACAGCCCTCGGTTACGGCTCTTCCACCATTCAGAGACAGATGTCACTTGGCC | 300 |
| Db | 241 | TTCAGCCACAGCCCTCGGTTACGGCTCTTCCACCATTCAGAGACAGATGTCACTTGGCC | 300 |

| | | | |
|----|------|--|------|
| Qy | 301 | CGCGCACGGAAACAACACTTTCAGTGTTCAGCTGCGACCTGGGGAGAACAGATGTACACTCGGCTC | 360 |
| Db | 301 | CGCGCACGGAAACAACACTTTCAGTGTTCAGCTGCGACCTGGGGAGAACAGATGTACACTCGGCTC | 360 |
| Qy | 361 | CGATGCACAGGCGAGCTCTGAGCTCCACACCTCCGACCGACGACCGACATGTACACTCGAGCC | 420 |
| Db | 361 | CGATGCACAGGCGAGCTCTGAGCTCCACACCTCCGACCGACGACCGACATGTACACTCGAGCC | 420 |
| Qy | 421 | CCGAGCAACAAGCAGACCCCGAGGCTCCACACCGCCCCCGAGCCACACGGATGTACACTCGAGCC | 480 |
| Db | 421 | CCGAGCAACAAGCAGACCCCGAGGCTCCACACCGCCCCCGAGCCACACGGATGTACACTCGAGCC | 480 |
| Qy | 481 | CCGAGCACACAGGCGCGCCCGCGAGGTCCACCGGC----- | 513 |
| Db | 481 | CCGAGCACACAGGCGCGCCCGCGAGGTCCACCGGC----- | 513 |
| Qy | 514 | -----CCGCGCGCCACCGATGTACACTCGAGCC | 540 |
| Db | 541 | CCGAGCACACAGGCGCGCCCGCGAGGTCCACACCGCGACCCACACGGATGTACACTCGAGCC | 600 |
| Qy | 541 | CCGAGCACACAGGCGCGCCCGCGAGGTCCACACCGCGACCCACACGGATGTACACTCGAGCC | 600 |
| Db | 601 | CCGAGCACACAGGCGCGCCCGCGAGGTCCACACCGCGACCCACACGGATGTACACTCGAGCC | 660 |
| Qy | 601 | CCGAGCACACAGGCGCGCTTGGGCTCCACCGCCCTTCAGATCCACAATGTACACTCGAGCC | 660 |
| Db | 661 | CCGAGCACACAGGCGCGCTTGGGCTCCACCGCCCTTCAGATCCACAATGTACACTCGAGCC | 720 |
| Qy | 661 | TCAGGCTCTGCATCAGGCTCCAGCTTCTACTCTGTGACACAACCGACCTCTGCCAGGACT | 720 |
| Db | 721 | TCAGGCTCTGCATCAGGCTCCAGCTTCTACTCTGTGACACAACCGACCTCTGCCAGGACT | 780 |
| Qy | 721 | ACCAACAACCCAGCAGCAGACAAGGACCTCCACCCAGACATTTCCAGCCACAACCTCGATACCT | 780 |
| Db | 781 | ACCAACAACCCAGCAGCAGACAAGGACCTCCATCTTCATTTCCAGCCACAACCTCGATACCT | 840 |
| Qy | 781 | CCTACACACCTTGGCAGGCGCATAGACAACAAGACTGATGCGCATGTAGACTCCACATAGACG | 840 |
| Db | 841 | CCTACACACCTTGGCAGGCGCATAGACAACAAGACTGATGCGCATGTAGACTCCACATAGACG | 900 |
| Qy | 841 | GTACCTCTCTCACTCCTCTCCAAATCAACAGACTTCTCCCAAGTTGCTACTCGGAGTCTCT | 900 |
| Db | 901 | GTACCTCTCTCACTCCTCTCCAAATCAACAGACTTCTCCCAAGTTGCTACTCGGAGTCTCT | 960 |
| Qy | 901 | TTCTTTTTCCTGTCTTTTTCATTTTCAATTTAAACCTCCAGTTTAAATCCCTCGGAAATATCC | 960 |
| Db | 961 | TTCTTTTTCCTGTCTTTTTCATTTTCAATTTAAACCTCCAGTTTAAATCCCTCGTCTGAAATATCC | 1020 |
| Qy | 961 | AGCACCACTACACCAAGAGCTGCGAAGACACTTCTGAAATGTTTTTGCAGATTAT | 1020 |
| Db | 1021 | AGCACCACTACACCAAGAGCTGCGAAGACACTTCTGAAATGTTTTTGCAGATTAT | 1080 |
| Qy | 1021 | AAACAAGGAGGAGTTTCTGGGCTCTCCAAATATTAAGTTCAAGGCGAGATCTGTGGTGGTA | 1080 |
| Db | 1081 | AAACAAGGAGGAGTTTCTGGGCTCTCCAAATATTAAGTTCAAGGCGAGATCTGTGGTGGTA | 1140 |
| Qy | 1081 | CAATTGACTGTGGCCTTCCAGAAAGGTACCAATCAATGTCCACGAGCGTGGAGAACACAGTTCC | 1140 |
| Db | 1141 | CAATTGACTGTGGCCTTCCAGAAAGGTACCAATCAATGTCCACGAGCGTGGAGAACACAGTTCC | 1200 |
| Qy | 1141 | AATCAAGTATTAACCGAAGCAGCTCTGCATTTAACTTGACGATCTCAGACTGACGGTGG | 1200 |
| Db | 1201 | AATCAAGTATTAACCGAAGCAGCTCTGCATTTAACTTGACGATCTCAGACTGACGGTGG | 1260 |
| Qy | 1201 | AGTCATGTGCAATTTCTTTTCTGCGCACAGTCTGGAGGCTGGAGGTCCAGGCTGGAGGATTC | 1260 |
| Db | 1261 | AGTCATGTGCAATTTCTTTTCTGCGCACAGTCTGGAGGCTGGAGGTCCAGGCTGGAGGATTC | 1320 |
| Qy | 1261 | GGGCGGCGTGGTGGTGTCTGTGTTTCTGTGTTGGTGGCGCATGTCTATCTCATTTGCGCTTG | 1320 |
| Db | 1321 | GGGCGGCGTGGTGGTGTCTGTGTTTCTGTGTTGGTGGCGCATGTCTATCTCATTTGCGCTTG | 1380 |
| Qy | 1321 | GCTGTCTGTCACTGCGCGCAAAAGAACTACGAGGACGTGTGACACTTTTTCAGCCCGGAGT | 1380 |

Db 1381 GCTGTCTGTAGTGGCCGCGGAAAGACTACGAGGAGCTGAGACATCTTTCAGCCGGAGAT 1440
Qy 1381 ACCTACATCTCTATAGAGGAGATACCCCACTACACACCCATGGAGGCTATATGCCCCCT 1440
Db 1441 ACCTACATCTCTATAGAGGAGATACCCCACTACACACCCATGGAGGCTATATGCCCCCT 1500
Qy 1441 AGCAGTACCGATGTAGGCGCCCTATGAGAAGGTTTTCGAGAGTAATGATGGAGCAGACCTC 1500
Db 1501 AGCAGTACCGATGTAGGCGCCCTATGAGAAGGTTTTCGAGAGTAATGATGGAGCAGACCTC 1560
Qy 1501 TCTTACCAAAACCCGACGATGGCAGCCACTTCTGCAACTTGTAGGGGACGTCGCCCTC 1560
Db 1561 TCTTACCAAAACCCGACGATGGCAGCCACTTCTGCAACTTGTAGGGGACGTCGCCCTC 1620
Qy 1561 TGAAGTGAAGTG 1572
Db 1621 TGAAGTGAAGTG 1632

RESULT 6
US-09-967-768A-224
Sequence 224 Application US/09967768A
GENERAL INFORMATION:
APPLICANT: Augmentus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-72
CURRENT APPLICATION NUMBER: US/09/967,768A
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236,109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,111
NUMBER OF SEQ ID NOS: 325
SOFTWARE: PatentIn version 3.0
SEQ ID NO 224
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-224

Query Match 94.2%; Score 1481.2; DB 36; Length 1721;
Best Local Similarity 95.5%; Pred. No. 5.5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

Qy 1 GAATTCCTGAGCTGTTGAATCTGTTGCCCCCTCCCAACCATTTTCAACCAACCATG 60
Db 1 GAATTCCTGAGCTGTTGAATCTGTTGCCCCCTCCCAACCATTTTCAACCAACCATG 60
Qy 61 ACACGGGAGCCAGATCTCTCTTCTCTCTGCTGCTGCTCTCTCAAGTGTCAAGTTGT 120
Db 61 ACACGGGAGCCAGATCTCTCTTCTCTCTGCTGCTGCTCTCTCAAGTGTCAAGTTGT 120
Qy 121 ACAGATTCGTGTATGCAAGCTCTACCCAGATGGAGAAAGAGAACTTGGGCTACCGAG 180
Db 121 ACAGATTCGTGTATGCAAGCTCTACCCAGATGGAGAAAGAGAACTTGGGCTACCGAG 180
Qy 181 AGAAGTTCATGTCCTGAGCTCTATGAGAAAGATGCTGTAGATATCAAGAGGATCTC 240
Db 181 AGAAGTTCATGTCCTGAGCTCTATGAGAAAGATGCTGTAGATATCAAGAGGATCTC 240
Qy 241 TCCAGGACAGCCCGGTTGAGGCTCTCCACCACTCAGGAGACAGATGTCACTTGGCC 300
Db 241 TCCAGGACAGCCCGGTTGAGGCTCTCCACCACTCAGGAGACAGATGTCACTTGGCC 300
Qy 301 CCGGACACGGAACCAAGCTTCAAGTTCAGCTGACCACTGGGAGACAGATGTCACTGAGTC 360
Db 301 CCGGACACGGAACCAAGCTTCAAGTTCAGCTGACCACTGGGAGACAGATGTCACTGAGTC 360
Qy 361 CCAATCACACAGGACCTGAGGCTTCACCAACCCCGCAGCCCAAGATGTCACTGAGCC 420

Db 361 CCAATCACACAGGACCTGAGGCTTCACCAACCCCGCAGCCCAAGATGTCACTGAGCC 420
Qy 421 CCGAGCAACAGCCAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 480
Db 421 CCGAGCAACAGCCAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 480
Qy 481 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 513
Db 481 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 540
Qy 514 -----CCGAGCCCAAGATGTCACTGAGCC 540
Db 541 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 600
Qy 541 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 600
Db 601 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 660
Qy 601 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 660
Db 661 CCGAGCAACAGCCCGGAGCTTCAACCGGCCCCCGGACCAAGATGTCACTGAGCC 720
Qy 661 TCAAGCTCTGATAGAGCTCAAGTCTTCTGATGAGCAAGGACCTTGGAGGAGCT 720
Db 721 TCAAGCTCTGATAGAGCTCAAGTCTTCTGATGAGCAAGGACCTTGGAGGAGCT 780
Qy 721 ACCAACAACCCAGCAGCAAGACATCCACCAAGATTCACCACTCTGATAT 780
Db 781 ACCAACAACCCAGCAGCAAGACATCCACCAAGATTCACCACTCTGATAT 840
Qy 781 CCTACCACTCTGAGCAGCTATAGACCAAGATGATGAGTATGAGTATGAGTATGAGC 840
Db 841 CCTACCACTCTGAGCAGCTATAGACCAAGATGATGAGTATGAGTATGAGTATGAGC 900
Qy 841 GATACCTCTGATAGCTCTCTCAATCAAGACCTTCCAGATGTCTAGTGGAGTCT 900
Db 901 GATACCTCTGATAGCTCTCTCAATCAAGACCTTCCAGATGTCTAGTGGAGTCT 960
Qy 901 TTTCTTTTCTGTCCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCC 960
Db 961 TTTCTTTTCTGTCCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCC 1020
Qy 961 AGCAGCACTACTACCAAGAGCTGAGAGACATTTCTGAATGTTTTTGGAGATTTAT 1020
Db 1021 AGCAGCACTACTACCAAGAGCTGAGAGACATTTCTGAATGTTTTTGGAGATTTAT 1080
Qy 1021 AAAAGAGGGGTTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGACAGATCTGTGGTA 1080
Db 1081 AAAAGAGGGGTTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGACAGATCTGTGGTA 1140
Qy 1081 CAATGACTTGGGCTTTCGAGAGGATCAATCAATGTCCAGACGTGGAGACACATTC 1140
Db 1141 CAATGACTTGGGCTTTCGAGAGGATCAATCAATGTCCAGACGTGGAGACACATTC 1200
Qy 1141 AATCAGTATTAAGAGAGAGAGCTCTCGATTAACCTGACATCTCAGACGTGACGTG 1200
Db 1201 AATCAGTATTAAGAGAGAGAGCTCTCGATTAACCTGACATCTCAGACGTGACGTG 1260
Qy 1201 AGTCAATGAGATTTCTTTCTCTGAGAGTGGAGCTGGAGCTGGAGCTGGAGCTG 1260
Db 1261 AGTCAATGAGATTTCTTTCTCTGAGAGTGGAGCTGGAGCTGGAGCTGGAGCTG 1320
Qy 1261 GCGCTGAGTGTGCTGT 1320
Db 1321 GCGCTGAGTGTGCTGT 1380
Qy 1321 GCTGTCTGTAGTGTGCGCGAAGAACTACGAGGACAGTGTCACTTTCACGCGGAGT 1380
Db 1381 GCTGTCTGTAGTGTGCGCGAAGAACTACGAGGACAGTGTCACTTTCACGCGGAGT 1440
Qy 1381 ACCTACATCTCTATGAGAGGATACCCCACTTACCAACCCATGAGGAGTGTGCCCCCT 1440

Db 1441 ACCTACATCTTATGACGAGTACCCCACTTACCAACCCATGGGCGCTATGTGCCCCCT 1500
Qy 1441 AGCAGTACCATCTTATGACGAGTACCCCTATGAGAGGTTTCTGACAGTAAATGTTGGAGAGCTTC 1500
Db 1501 AGCAGTACCATCTTATGACGAGTACCCCTATGAGAGGTTTCTGACAGTAAATGTTGGAGAGCTTC 1560
Qy 1501 TCTTACACAAACCCAGAGTGGAGGCACTTCTGCAACTTGTAGGGGAGCTGGCCCTC 1560
Db 1561 TCTTACACAAACCCAGAGTGGAGGCACTTCTGCAACTTGTAGGGGAGCTGGCCCTC 1620
Qy 1561 TGAGCTGAGTGG 1572
Db 1621 TGAGCTGAGTGG 1632

RESULT 7
US-10-007-926A-58
; Sequence 58, Application US/10007926A
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOUIGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NEUBEN, CATHERINE
; APPLICANT: VIKERS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 58
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mucin 1, transmembrane (MUC1) gene.
US-10-007-926A-58

Query Match 94.2%; Score 1481.2; DB 38; Length 1721;
Best Local Similarity 95.5%; Pred. No. 5.5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

Qy 1 GAATTCCTGGCTGCTGATCTGTTCTGCCCCCTCCCAACCCATTTCACACACACCATG 60
Db 1 GAATTCCTGGCTGCTGATCTGTTCTGCCCCCTCCCAACCCATTTCACACACACCATG 60
Qy 61 ACACCGGGGACCCGAGTCTCTCTTCTGCTGCTGCTCTCTCAAGTGTTCAGTTGTT 120
Db 61 ACACCGGGGACCCGAGTCTCTCTTCTGCTGCTGCTCTCTCTCAAGTGTTCAGTTGTT 120
Qy 121 ACAGGTTCTGTCATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGAGTACCCAG 180
Db 121 ACAGGTTCTGTCATGCAAGCTTACCCAGAGTGAAGAAAGAGACTTCGAGTACCCAG 180
Qy 181 AGAAGTTCAAGTCCAGCTCTACTGAGAAAGATGTCGTGATGTCAGAGAGGTCATC 240
Db 181 AGAAGTTCAAGTCCAGCTCTACTGAGAAAGATGTCGTGATGTCAGAGAGGTCATC 240
Qy 241 TCCAGCCACAGCCCGGTTGAGGCTCTCTCACTCACTCAAGGAGACAGATGTCACTTG 300
Db 241 TCCAGCCACAGCCCGGTTGAGGCTCTCTCACTCACTCAAGGAGACAGATGTCACTTG 300
Qy 301 CCAGGCAAGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTCACTTGAGT 360
Db 301 CCAGGCAAGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTCACTTGAGT 360
Qy 361 CCAGTCAACAGAGGCTTGGGCTCACTCAACCCCGGAGCCAGATGTCACTTGAGG 420
Db 361 CCAGTCAACAGAGGCTTGGGCTCACTCAACCCCGGAGCCAGATGTCACTTGAGG 420

Qy 421 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGCCCGCCCGGACCAAGATGTCACTTGAGG 480
Db 421 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGCCCGCCCGGACCAAGATGTCACTTGAGG 480
Qy 481 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 513
Db 481 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 540
Qy 514 -----CCGCGCGCCACCGGTCACTTGAGG 540
Db 541 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 600
Qy 541 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 600
Db 601 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 660
Qy 601 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 660
Db 661 CCGGACAAAGAGCCAGCCCGGAGCTCCACCGGCTCCACCGGCTCCACCGGCTCCACCGG 720
Qy 661 TCAAGGCTCTGATCAAGGCTCAAGCTTCTACTCTGTGTCACAAAGGACCTTCTCCAGG 780
Db 721 TCAAGGCTCTGATCAAGGCTCAAGCTTCTACTCTGTGTCACAAAGGACCTTCTCCAGG 780
Qy 721 ACCAACAACCCAGCAGACAGCACTCCACCAAGCACTTCCACCAAGCACTTCTGATAT 780
Db 781 ACCAACAACCCAGCAGACAGCACTCCACCAAGCACTTCCACCAAGCACTTCTGATAT 840
Qy 781 CTTACACACCTTGCAGCAGATTAAGCAAGATGATGAGTGAAGTGAAGTGAAGTGAAGTGA 840
Db 841 CTTACACACCTTGCAGCAGATTAAGCAAGATGATGAGTGAAGTGAAGTGAAGTGAAGTGA 900
Qy 841 GTACCTCTCTACCTCTCTCAATCAAGACACTTCTCCAGTGTCTAAGGAGTCTCT 900
Db 901 GTACCTCTCTACCTCTCTCAATCAAGACACTTCTCCAGTGTCTAAGGAGTCTCT 960
Qy 901 TTTCTTTTCTGTCTTTTCAATTTCAAACTCCAGTTTAACTCTCTCTGAGAGATCC 960
Db 961 TTTCTTTTCTGTCTTTTCAATTTCAAACTCCAGTTTAACTCTCTCTGAGAGATCC 1020
Qy 961 AGCAGCACTACTCAAGAGAGCTGACAGAGACATTTCTGAAATGTTTGGAGATTTAT 1020
Db 1021 AGCAGCACTACTCAAGAGAGCTGACAGAGACATTTCTGAAATGTTTGGAGATTTAT 1080
Qy 1021 AAACAGAGGGGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1080
Db 1081 AAACAGAGGGGTTTCTGGGCTCTCCAAATTAAGTTCAAGGCAAGATCTGTGGTGA 1140
Qy 1081 CAATTGACTGTGGCTTCCGAGAGTACCAATGTCACAGCTGAGACAGATTC 1140
Db 1141 CAATTGACTGTGGCTTCCGAGAGTACCAATGTCACAGCTGAGACAGATTC 1200
Qy 1141 AATCAGTATTAAGAGAGAGAGCTCTCGATTAACCTGACAGATCTCAGACCTCAGCTG 1200
Db 1201 AATCAGTATTAAGAGAGAGAGCTCTCGATTAACCTGACAGATCTCAGACCTCAGCTG 1260
Qy 1201 AGTCATGAGCATTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1261 AGTCATGAGCATTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Qy 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1321 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Qy 1321 GCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Db 1381 GCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Qy 1381 ACCATCATCTATGAGAGTACCCCACTTACCAACCCAGGAGCTGATGTGCCCCCT 1440
Db 1441 ACCTACATCTTATGAGAGTACCCCACTTACCAACCCAGGAGCTGATGTGCCCCCT 1500

```
QY 1441 AGCAGTACCGATGCTAGCCCTTATGAGAAGTTTTCGACGTAATGTTGGCAGAGCCTTC 1500
DB 1501 AGCAGTACCGATGCTAGCCCTTATGAGAAGTTTTCGACGTAATGTTGGCAGAGCCTTC 1560
QY 1501 TCTTACAAACCCAGCAGATGAGCAGCACTTCTGCCAATTGTAGAGGAGCAGTGCCTTC 1560
DB 1561 TCTTACAAACCCAGCAGATGAGCAGCACTTCTGCCAATTGTAGAGGAGCAGTGCCTTC 1620
QY 1561 TGAGCTGAGTGG 1572
DB 1621 TGAGCTGAGTGG 1632

RESULT 8
US-10-029-517-3
; Sequence 3, Application US/10029517
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan J. Myers
; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION
; FILE REFERENCE: RTS-0352
; CURRENT APPLICATION NUMBER: US/10/029,517
; NUMBER OF SEQ ID NOS: 107
; SEQ ID NO 3
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (58)...(1605)
US-10-029-517-3

Query Match 94.2%; Score 1481.2; DB 38; Length 1721;
Best Local Similarity 95.5%; Pred. No. 5.5e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 1 GAATTCCTGCGCTGCTGATGCTGTCGCCCCCTCCGACCCATTTCACCAACCAAG 60
DB 1 GAATTCCTGCGCTGCTGATGCTGTCGCCCCCTCCGACCCATTTCACCAACCAAG 60
QY 61 ACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTGTTCAGTTGTT 120
DB 61 ACACCGGGGACCCAGTCTCTCTTCTCTCTGCTGCTCTCTCAAGTGTTCAGTTGTT 120
QY 121 ACAGGTTCTGTCATGCAAGCTCTACCCGAGTGAAGAAAGAGACTTGGCTACCCAG 180
DB 121 ACAGGTTCTGTCATGCAAGCTCTACCCGAGTGAAGAAAGAGACTTGGCTACCCAG 180
QY 181 AGAAGTTCAGTCCAGCTCTAAGAGAAAGATGCTGTGATGACACAGAGCTATCTC 240
DB 181 AGAAGTTCAGTCCAGCTCTAAGAGAAAGATGCTGTGATGACACAGAGCTATCTC 240
QY 241 TCCAGCCACAGCCCGGTTCAAGCTCTCTCAACAATCAGGAGACAGAGATGCTAGCTC 300
DB 241 TCCAGCCACAGCCCGGTTCAAGCTCTCTCAACAATCAGGAGACAGAGATGCTAGCTC 300
QY 301 CCGGCGACCGGAAACAGGTTCAAGTTCAGCTGACCACTGGGAGACAGAGATGCTAGCTC 360
DB 301 CCGGCGACCGGAAACAGGTTCAAGTTCAGCTGACCACTGGGAGACAGAGATGCTAGCTC 360
QY 361 CCACTGACAGGCGCAGCTTCCAGCTCTCAACAACCCGCGACAGCCAGATGCTAGCTC 420
DB 361 CCACTGACAGGCGCAGCTTCCAGCTCTCAACAACCCGCGACAGCCAGATGCTAGCTC 420
QY 421 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 480
DB 421 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 480
QY 481 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 513
DB 481 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 540
```

```
QY 514 -----CCGCGCGCCACAGGTTTCACTGAGCC 540
DB 541 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 600
QY 541 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 600
DB 601 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 660
QY 601 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 660
DB 661 CCGGACAAACAGCAGCAGCCCGGCTTCAACCGCCCGCCGACAGGATGCTAGCTC 720
QY 661 TCAAGCTCTGCAATCAGGCTCAGCTTCTGCTGTCACAAAGCAGCTTCCGAGGCT 720
DB 721 TCAAGCTCTGCAATCAGGCTCAGCTTCTGCTGTCACAAAGCAGCTTCCGAGGCT 780
QY 721 ACCAACAACCCAGCAGCAGGACCTCCAGGATTCGAGCAGCACTCTGATACT 780
DB 781 ACCAACAACCCAGCAGCAGGACCTCCAGGATTCGAGCAGCACTCTGATACT 840
QY 781 CTTACCAACCTTGGCAGCCATGACCAAGACTGATGCGACTCAGCATAGCAGC 840
DB 841 CTTACCAACCTTGGCAGCCATGACCAAGACTGATGCGACTCAGCATAGCAGC 900
QY 841 GTACTCTCTGCACTCTCTCAATCAGCACTTCTCCAGTTGCTACTGAGGCTCT 900
DB 901 GTACTCTCTGCACTCTCTCAATCAGCACTTCTCCAGTTGCTACTGAGGCTCT 960
QY 901 TTTCTTTTCTGCTCTTCTTCAATTCATTAACCTCAGTTTAACTCTCTGAGAAATCCC 960
DB 961 TTTCTTTTCTGCTCTTCTTCAATTCATTAACCTCAGTTTAACTCTCTGAGAAATCCC 1020
QY 961 AGCAGCAGTACTACCAAGAGCTCAGAGAGACATTTTGAATGTTTTTTCAGATTTAT 1020
DB 1021 AGCAGCAGTACTACCAAGAGCTCAGAGAGACATTTTGAATGTTTTTTCAGATTTAT 1080
QY 1021 AAAAAGAGGGGTTTTCTGGGCTCTCCAAATATTAAGTTCAGGCGCAGGTCGAGTGA 1080
DB 1081 AAAAAGAGGGGTTTTCTGGGCTCTCCAAATATTAAGTTCAGGCGCAGGTCGAGTGA 1140
QY 1081 CAATTGACTCTGAGCTTCCGAGAAAGTACATCATATGTCACAGCTGAGAGACAGTTC 1140
DB 1141 CAATTGACTCTGAGCTTCCGAGAAAGTACATCATATGTCACAGCTGAGAGACAGTTC 1200
QY 1141 AATCAGTATPAAACGGAAGCAGCTCTGATATTAACCTGACGATCTCAGAGCTG 1200
DB 1201 AATCAGTATPAAACGGAAGCAGCTCTGATATTAACCTGACGATCTCAGAGCTG 1260
QY 1201 AGTATGTCGCAATTTCTTTCTCTGCCCCAGTCTGGGGCTGGGGTCCAGGCTGGGATTC 1260
DB 1261 AGTATGTCGCAATTTCTTTCTCTGCCCCAGTCTGGGGCTGGGGTCCAGGCTGGGATTC 1320
QY 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1321 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1321 GCTGTCGTCAGGTCGCGCGGAAAGAACTAAGGAGAGCTGGAACATTTTCCAGCCCGGAA 1380
DB 1381 GCTGTCGTCAGGTCGCGCGGAAAGAACTAAGGAGAGCTGGAACATTTTCCAGCCCGGAA 1440
QY 1381 ACCTACCATCTATGAGGAGATCCCACTTACACACCACTGAGGCGCTATGTCCTCT 1440
DB 1441 ACCTACCATCTATGAGGAGATCCCACTTACACACCACTGAGGCGCTATGTCCTCT 1500
QY 1441 AGCAGTACCGATGCTAGCCCTTATGAGAAGTTTTCGACGTAATGTTGGCAGAGCCTTC 1500
DB 1501 AGCAGTACCGATGCTAGCCCTTATGAGAAGTTTTCGACGTAATGTTGGCAGAGCCTTC 1560
QY 1501 TCTTACAAACCCAGCAGTGAAGCTGACCACTTCTGCCAATTGTAGAGGAGCAGTGCCTTC 1560
DB 1561 TCTTACAAACCCAGCAGTGAAGCTGACCACTTCTGCCAATTGTAGAGGAGCAGTGCCTTC 1620
QY 1561 TGAGCTGAGTGG 1572
```


Db 1621 TGAGCTGAGTGG 1632

RESULT 9
US-10-097-340-211

```

Sequence 211, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumei ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For the Identification,
Title of Invention: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 211
LENGTH: 1721
TYPE: DNA
ORGANISM: Homo sapiens
US-10-097-340-211

```

| | | | | |
|----------------------------|--------|---------------------|------------|--------------|
| Query Match | 94.2%; | Score 1481.2; | DB 39; | Length 1721; |
| Best Local Similarity | 95.5%; | Pred. No. 5.5e-280; | | |
| Matches 1559; Conservative | 0; | Mismatches 13; | Indels 60; | Gaps 1 |

| | | | |
|----|-----|--|-----|
| Qy | 1 | GATTTCCCTGGCTGCTGAAATGTGTCGCCCCCTCCACCAATTACACACACATG | 60 |
| Db | 1 | GAAATTCCTGGCTGCTGAAATCTGTTCGCCCCCTCCACCAATTACACACACATG | 60 |
| Qy | 61 | ACACGGGACCCAGCTCTCCTTTCTTCTGTGCTGCTCTCAAGCTTACAGTTT | 120 |
| Db | 61 | ACACGGGGACCCAGCTCTCCTTTCTTCTGTGCTGCTCTCAAGCTTACAGTTT | 120 |
| Qy | 121 | ACAGGTTCTGTCATGCAAGCTCTACCCACAGTGGAGAAAGAGACTTGGGTACCG | 180 |
| Db | 121 | ACAGGTTCTGTCATGCAAGCTCTACCCACAGTGGAGAAAGAGACTTGGGTACCG | 180 |
| Qy | 181 | AGAAAGTTACAGGCCCGAGCTCTACAGAAAGATGCTGTAAGATGACACAGCGTATC | 240 |
| Db | 181 | AGAAAGTTACAGGCCCGAGCTCTACAGAAAGATGCTGTAAGATGACACAGCGTATC | 240 |
| Qy | 241 | TCGAGTCACAGCCCTCCGATTCAAGCTCTTCACACTCAAGGACAGAAATCACTTGGCC | 300 |

[illegible]

QY 1321 GCTGTCTGTAGTCCGCGGAAAGAACTACCGGAGCTGAGATCTTTTCAGCCCGGGAT 1380
 DB 1381 GCTGTCTGTAGTCCGCGGAAAGAACTACCGGAGCTGAGATCTTTTCAGCCCGGGAT 1440
 QY 1381 ACCACCACTCTATGAGAGATACCCCACTACCAACCCATGGGGGCTATGAGCCCTC 1440
 DB 1441 ACCACCACTCTATGAGAGATACCCCACTACCAACCCATGGGGGCTATGAGCCCTC 1500
 QY 1441 AGCGATCCGATGTGAGCCCTATATGAAAGTTTCTGCAAGTAAATGTTGCGAGAGCTTC 1500
 DB 1501 AGCGATCCGATGTGAGCCCTATATGAAAGTTTCTGCAAGTAAATGTTGCGAGAGCTTC 1560
 QY 1501 TCTTACCAAAACCCAGACAGTGCACCTTCTGCACTTGTAGAGGCGACGTCGCTTC 1560
 DB 1561 TCTTACCAAAACCCAGACAGTGCACCTTCTGCACTTGTAGAGGCGACGTCGCTTC 1620
 QY 1561 TGAGCTAGTGG 1572
 DB 1621 TGAGCTAGTGG 1632

RESULT 10
 US-10-171-311-155
 ; Sequence 155, Application US/10171311
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlegel, Robert
 ; APPLICANT: Chen, Yan
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Monahan, John
 ; APPLICANT: Kamelkar, Shubhangi
 ; APPLICANT: Glatz, Karen
 ; APPLICANT: Gannavarapu, Manjula
 ; APPLICANT: Hoerish, Sebastian
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
 ; FILE REFERENCE: OF CERVICAL CANCER
 ; FILE REFERENCE: MRI-035
 ; CURRENT APPLICATION NUMBER: US/10/171,311
 ; CURRENT FILING DATE: 2002-06-12
 ; PRIOR APPLICATION NUMBER: US 60/298,159
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,155
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/335,936
 ; PRIOR FILING DATE: 2001-11-14
 ; NUMBER OF SEQ ID NOS: 238
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 155
 ; LENGTH: 1721
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-171-311-155

Query Match 94.2%; Score 1481.2; DB 41; Length 1721;
 Best Local Similarity 95.5%; Pred. No. 5.5e-280;
 Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 1 GAATTCCTGCTGCTTGAATCTGTTGCCCCCTCCCAACCATTTTACCAACCATG 60
 DB 1 GAATTCCTGCTGCTTGAATCTGTTGCCCCCTCCCAACCATTTTACCAACCATG 60
 QY 61 AACCGGAGCAACGAGTCTCTTCTTCTGCTGCTCTCTCAAGTGTCTTACAGTTGTT 120
 DB 61 AACCGGAGCAACGAGTCTCTTCTTCTGCTGCTCTCTCAAGTGTCTTACAGTTGTT 120
 QY 121 ACAGGTTCTGATGATGAGCTCTACCCAGGTGGAAGAAAGAGACTTGGCTACCCAG 180
 DB 121 ACAGGTTCTGATGATGAGCTCTACCCAGGTGGAAGAAAGAGACTTGGCTACCCAG 180
 QY 181 AGAAGTTCAAGTCCGAGCTCTACTGAAAGAAAGATGCTGTGATGACAGAGGTTATC 240
 DB 181 AGAAGTTCAAGTCCGAGCTCTACTGAAAGAAAGATGCTGTGATGACAGAGGTTATC 240

QY 241 TCAGCCACAGCCCGGTTTCAGGCTCTCTCCACCACTACAGGAGACAGATGTCACTTGGCC 300
 DB 241 TCAGCCACAGCCCGGTTTCAGGCTCTCTCCACCACTACAGGAGACAGATGTCACTTGGCC 300
 QY 301 CCGGACCAAGGAGACAGCTTCAAGTTGAGCTGACCACTGGGAGACAGATGTCACTTGGCC 360
 DB 301 CCGGACCAAGGAGACAGCTTCAAGTTGAGCTGACCACTGGGAGACAGATGTCACTTGGCC 360
 QY 361 CCAATCACTCAAGGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 420
 DB 361 CCAATCACTCAAGGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 420
 QY 421 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 480
 DB 421 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 480
 QY 481 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 513
 DB 481 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 513
 QY 514 -----CCGCGCGCCAGGTTTCACTTGGCC 540
 DB 514 -----CCGCGCGCCAGGTTTCACTTGGCC 540
 QY 541 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 600
 DB 541 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 600
 QY 541 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 600
 DB 541 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 600
 QY 601 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 660
 DB 601 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 660
 QY 661 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 720
 DB 661 CCGGACCAAGGAGACAGCTTGGCTTCCACACCCGCGCACATGTCACTTCACTGACC 720
 QY 721 TCAGGCTCTGATCAAGGCTCACTTCTACTGCTGTCACAAAGGAGACCTGTCAGAGGCT 780
 DB 721 TCAGGCTCTGATCAAGGCTCACTTCTACTGCTGTCACAAAGGAGACCTGTCAGAGGCT 780
 QY 781 AACCAACCCGAGGAGACAGCTTCCACACCCGCGCACATGTCACTTCACTGACC 840
 DB 781 AACCAACCCGAGGAGACAGCTTCCACACCCGCGCACATGTCACTTCACTGACC 840
 QY 841 CCAACCACTTCCAGGAGACAGCTTCCACACCCGCGCACATGTCACTTCACTGACC 900
 DB 841 CCAACCACTTCCAGGAGACAGCTTCCACACCCGCGCACATGTCACTTCACTGACC 900
 QY 901 GTACTCTCTTCACTCTCTCTCAATCAAGGACCTTCCAGGAGACAGCTTCACTGACC 960
 DB 901 GTACTCTCTTCACTCTCTCTCAATCAAGGACCTTCCAGGAGACAGCTTCACTGACC 960
 QY 961 TTTCTTTTCTGCTTCTTCAATTTCAAACTTCAAGTTTATTTCTCTCTGAAAGATCCC 1020
 DB 961 TTTCTTTTCTGCTTCTTCAATTTCAAACTTCAAGTTTATTTCTCTCTGAAAGATCCC 1020
 QY 1021 AGCACCACTTACCAAGAGCTGACAGAGACATTTCTGAAGTTTTCAGATTTAT 1080
 DB 1021 AGCACCACTTACCAAGAGCTGACAGAGACATTTCTGAAGTTTTCAGATTTAT 1080
 QY 1081 AAACAGGAGGTTTCTGGGCTCTCAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1140
 DB 1081 AAACAGGAGGTTTCTGGGCTCTCAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1140
 QY 1141 CAATTGACTCTGAGCTTCCAGAAAGATTAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1200
 DB 1141 CAATTGACTCTGAGCTTCCAGAAAGATTAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1200
 QY 1201 AATCAATTAATTAAGGAGAGAGCTCTCAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1260
 DB 1201 AATCAATTAATTAAGGAGAGAGCTCTCAATTAATTAAGTTTTCAGAGGAGATCTGAGTGTGA 1260
 QY 1261 AGTATGTGCAATTTCTTCTCTGCGCAAGTCTGAGGAGTCTGAGGAGTCTGAGGAGATC 1320
 DB 1261 AGTATGTGCAATTTCTTCTCTGCGCAAGTCTGAGGAGTCTGAGGAGTCTGAGGAGATC 1320

Qy 1381 ACCGACATCCATGAGAGTACCCACCTACACACACGAGGAGGATGCCCCCT 1440
Db 1441 ACCGACATCCATGAGAGTACCCACCTACACACACGAGGAGGATGCCCCCT 1500
Qy 1441 AGCAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTGCAAGTATATGATGACAGCCTTC 1500
Db 1501 AGCAGTACCGATCGTAGCCCTTATGAGAAAGTTTCTGCAAGTATATGATGACAGCCTTC 1560
Qy 1501 TCTTACACAAACCCAGCAGTGGAGCCACTTCTGCAACTTGTGAGGAGCAGTCGCCCTC 1560
Db 1561 TCTTACACAAACCCAGCAGTGGAGCCACTTCTGCAACTTGTGAGGAGCAGTCGCCCTC 1620
Qy 1561 TGAGCTGAGTGG 1572
Db 1621 TGAGCTGAGTGG 1632

RESULT 13
US-60-209-009-103
; Sequence 103, Application US/60209009
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027 P
; CURRENT APPLICATION NUMBER: US/60/209,009
; CURRENT FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO: 103
; LENGTH: 2678
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 235943.39c
US-60-209-009-103

Query Match 94.2%; Score 1481.2; DB 64; Length 2678;
Best Local Similarity 95.5%; Pred. No. 5.8e-280;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

Qy 1 GAATTCCTGGCTGCTGATCTGTCTGCCCCCTCCCACTTTCACACACCATG 60
Db 1 GAATTCCTGGCTGCTGATCTGTCTGCCCCCTCCCACTTTCACACACCATG 60
Qy 61 ACACCGGAGCACCAAGTCTCTTCTCTGCTGCTCTCTCAAGTCTTACAGTGT 120
Db 61 ACACCGGAGCACCAAGTCTCTTCTCTGCTGCTCTCTCAAGTCTTACAGTGT 120
Qy 121 ACAGTTTCTGCTGATGACAGCTCTACCCAGAGTGAAGAAAGAGACTTCGCTACCCAG 180
Db 121 ACAGTTTCTGCTGATGACAGCTCTACCCAGAGTGAAGAAAGAGACTTCGCTACCCAG 180
Qy 181 AGAAGTTCAAGTCCAGCTCTACTGAGAGATGTGTGATGATGACAGAGGTACTG 240
Db 181 AGAAGTTCAAGTCCAGCTCTACTGAGAGATGTGTGATGATGACAGAGGTACTG 240
Qy 241 TCCAGCACAAGCCCGGTTCAAGGCTCTCCACACTCAGGAGACAGATGTCACTCTGACC 300
Db 241 TCCAGCACAAGCCCGGTTCAAGGCTCTCCACACTCAGGAGACAGATGTCACTCTGACC 300
Qy 301 CCGGACACCGAAACAGCTTCAGGTTCAAGTCCAGCTTGGGAGACAGATGTCACTCTGACC 360
Db 301 CCGGACACCGAAACAGCTTCAGGTTCAAGTCCAGCTTGGGAGACAGATGTCACTCTGACC 360
Qy 361 CCACTACACAGGACGAGCTTGGGCTCCACACACCGGACAGCCACAGATGTCACTCTGACC 420
Db 361 CCACTACACAGGACGAGCTTGGGCTCCACACACCGGACAGCCACAGATGTCACTCTGACC 420
Qy 421 CCGGAAACAAAGCCAGCCCGGAGCTCCACCGGCCCCCGGACCAAGGTGTCACTCTGACC 480
Db 421 CCGGAAACAAAGCCAGCCCGGAGCTCCACCGGCCCCCGGACCAAGGTGTCACTCTGACC 480

Db 421 CCGGAAACAAAGCCAGCCCGGAGCTCCACCGGCCCCCGGACCAAGGTGTCACTCTGACC 480
Qy 481 CCGGAAACAAAGCCAGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 513
Db 481 CCGGAAACAAAGCCAGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 540
Qy 514 -----CCGCGGACCAAGGTGTCACTCTGACC 540
Db 541 CCGGACACAGGCGGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 600
Qy 541 CCGGACACAGGCGGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 600
Db 601 CCGGACACAGGCGGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 660
Qy 601 CCGGACACAGGCGGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 660
Db 661 CCGGACACAGGCGGCCCGGAGCTCCACCGGCCCCCGGAGCTCCACCGGCCCCCGGAGCT 720
Qy 661 TCAAGCTCTGATCAGGCTCAGCTTCTTACTCTGCTGACAAAGGACCTCTGACAGGCT 720
Db 721 TCAAGCTCTGATCAGGCTCAGCTTCTTACTCTGCTGACAAAGGACCTCTGACAGGCT 780
Qy 721 ACCACAAACCCAGCCAGCAAGACCTCCACCCAGCACTTCCACCCAGCACTCTGATAT 780
Db 781 ACCACAAACCCAGCCAGCAAGACCTCCACCCAGCACTTCCACCCAGCACTCTGATAT 840
Qy 781 CCTACACCCCTTGGCAGCAGTACACCAAGACTGATGACAGTACCTACCATAGACAG 840
Db 841 CCTACACCCCTTGGCAGCAGTACACCAAGACTGATGACAGTACCTACCATAGACAG 900
Qy 841 GTACCTCTCTCAGCTCTCTCAATACAGACTTCTCCGAGTGTCTACAGGAGTCT 900
Db 901 GTACCTCTCTCAGCTCTCTCAATACAGACTTCTCCGAGTGTCTACAGGAGTCT 960
Qy 901 TTTCTTTTCTGCTCTTTTACATTTTCAAACTCCAGTTTATCTCTCTGAGAGATCC 960
Db 961 TTTCTTTTCTGCTCTTTTACATTTTCAAACTCCAGTTTATCTCTCTGAGAGATCC 1020
Qy 961 AGCAGCAGTACTACCAAGAGCTGAGAGAGCATTTCTGAAATGTTTTTGGACATTTAT 1020
Db 1021 AGCAGCAGTACTACCAAGAGCTGAGAGAGCATTTCTGAAATGTTTTTGGACATTTAT 1080
Qy 1021 AAACAAGGAGGTTTTCTGAGGCTCTCCAAATTAATTAAGTTCAGGCAGAGTCTGTGTGTA 1080
Db 1081 AAACAAGGAGGTTTTCTGAGGCTCTCCAAATTAATTAAGTTCAGGCAGAGTCTGTGTGTA 1140
Qy 1081 CAATTGACTTGGCTTCCGAGAGGTACATCAATGTCCAGACCTGAGACACAGTTC 1140
Db 1141 CAATTGACTTGGCTTCCGAGAGGTACATCAATGTCCAGACCTGAGACACAGTTC 1200
Qy 1141 AATCAGTATAAAGAGAGAGGCTCTCGATATTAACCTGACATCTCAGACGTCAGGCTG 1200
Db 1201 AATCAGTATAAAGAGAGAGGCTCTCGATATTAACCTGACATCTCAGACGTCAGGCTG 1260
Qy 1201 AGTCATGAGCATTTCTCTTCTGTCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1260
Db 1261 AGTCATGAGCATTTCTCTTCTGTCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1320
Qy 1261 GCGCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1321 GCGCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Qy 1321 GCTGTCTGATGAGCGCCGAGAAAGAACTACCGGAGAGCTGAGCAATCTTTCAGCCGAGAT 1380
Db 1381 GCTGTCTGATGAGCGCCGAGAAAGAACTACCGGAGAGCTGAGCAATCTTTCAGCCGAGAT 1440
Qy 1381 ACCTACATCTATGAGAGGTACCCCACTTACACACCACTAGGAGGAGTCTGCTGCT 1440
Db 1441 ACCTACATCTATGAGAGGTACCCCACTTACACACCACTAGGAGGAGTCTGCTGCT 1500
Qy 1441 AGCAGTACCGATCTGAGCCCTTATGAGAGGTTCTGAGTAAATGATGAGAGGCTG 1500
Db 1501 AGCAGTACCGATCTGAGCCCTTATGAGAGGTTCTGAGTAAATGATGAGAGGCTG 1560

Thu May. 8 18:53:09 2003

us-09-658-621b-1.rpm

Page 15

| Accession | Sequence | Length |
|-----------|---|--------|
| QY 1501 | TCCTTACACAAACCCAGCAGTGGCAGCCATTCTGCCTGTAAGGGAGCAGTGCCTC | 1560 |
| Db 1561 | TCCTTACACAAACCCAGCAGTGGCAGCCATTCTGCCTGTAAGGGAGCAGTGCCTC | 1620 |
| QY 1561 | TGAGCTGAGTGG | 1572 |
| Db 1621 | TGAGCTGAGTGG | 1632 |

RESULT 14
US-09-716-473-2290/c

```

: Sequence 2290, Application US/09716473
: GENERAL INFORMATION:
: APPLICANT: Williamson, Mark
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2043-001
: CURRENT APPLICATION NUMBER: US/09/716,473
: CURRENT FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: 60/166,502
: PRIOR FILING DATE: 1999-11-19
: NUMBER OF SEQ ID NOS: 2933
: SOFTWARE: PaetSed for Windows Version 4.0
: SEQ ID NO 2290
: LENGTH: 2574
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-716-473--2290

```

| | | | | |
|----------------------------|--------|-------------------|------------|--------------|
| Query Match | 92.5%; | Score 1454.4; | DB 28; | Length 2574; |
| Best Local Similarity | 95.3%; | Pred. No. 1e-274; | | |
| Matches 1534; Conservative | 0; | Mismatches 16; | Indels 60; | Gaps 1 |

| | | | |
|----|------|--|------|
| QY | 1 | GAATTCCTGGCTGTAAATCTTTCTGCCCCCTCCACCAATTTACCAACACG | 60 |
| Db | 2572 | GAATTCCTGGCTGTAAATCTTTCTGCCCCCTCCACCAATTTACCAACACG | 2513 |
| QY | 61 | ACACCGGGACCCAGCTCTCTTTCTCTCTGCTCTCTCTCAAGAGCTTAAGTGT | 120 |
| Db | 2512 | ACACCGGGACCCAGCTCTCTTTCTCTCTGCTCTCTCTCAAGAGCTTAAGTGT | 2455 |
| QY | 121 | ACAGGTTCTGGTCAATGCAAGCTTACCCAGGTGGAGAAAGAGACTTGGCTACC | 180 |
| Db | 2452 | ACAGGTTCTGGTCAATGCAAGCTTACCCAGGTGGAGAAAGAGACTTGGCTACC | 2393 |
| QY | 181 | AGAAGTTCAATGCCAGGCTTACTAGAGAAATGCTGAGTATACACGAGGATAC | 240 |
| Db | 2392 | AGAAGTTCAATGCCAGGCTTACTAGAGAAATGCTGAGTATACACGAGGATAC | 2333 |
| QY | 241 | TCGACGCCACAGCCCGGGTTCAAGCTCTCTCCACATCTCAGAGGACAGAAATGTCACTCTGGCC | 300 |
| Db | 2332 | TCGACGCCACAGCCCGGGTTCAAGCTCTCTCCACATCTCAGAGGACAGAAATGTCACTCTGGCC | 2273 |
| QY | 301 | CCGGCCACGGAAACAGCTTCAAGTTCAAGCTTCAAGCTGGGGAACAGATGTCACTCTGGCTC | 360 |
| Db | 2272 | CCGGCCACGGAAACAGCTTCAAGTTCAAGCTTCAAGCTGGGGAACAGATGTCACTCTGGCTC | 2213 |
| QY | 361 | CGAATCAACAGGCGAGGCTCTGGGGCTCCACCAACCGGCGACGACCAAGATGTCACTCAAGCC | 420 |
| Db | 2212 | CGAATCAACAGGCGAGGCTCTGGGGCTCCACCAACCGGCGACGACCAAGATGTCACTCAAGCC | 2153 |
| QY | 421 | CCGGAACAACAGCGAGCCCGGGCTTCAACCGGCTCCCGGAGCCACAGTGTCACTCTGGCC | 480 |
| Db | 2152 | CCGGAACAACAGCGAGCCCGGGCTTCAACCGGCTCCCGGAGCCACAGTGTCACTCTGGCC | 2093 |
| QY | 481 | CCGGAACAACAGGCGGCGCCCGGGGCTTCAACCGGC----- | 513 |
| Db | 2092 | CCGGAACAACAGGCGGCGCCCGGGGCTTCAACCGGC----- | 2033 |
| QY | 514 | -----CCGGCGGCCACAGTGTCACTCTGGCC | 540 |

| | | | | |
|---|---|------|--|------|
| D | b | 2032 | CCGGAACACAGGCGCCGCCCGGAGCTCAACCGGCGCCGAGACCAAGGATGCACTCGGCG | 1973 |
| O | y | 541 | CCGGACACAGGCGCGCGCGCGGCTCAACCGCGCCCGCGGACCAAGGATGCACTCGGCG | 600 |
| D | b | 1972 | CCGGACACAGGCGCGCGCGGAGCTCAACCGCGCCCGGAGCCATGGATGCACTCGGCG | 1913 |
| O | y | 601 | CCGGACACAGGCGCGGCTTGGGCTCAACCGCGCCCTCGAGTCAAGATGCACTCGGCG | 660 |
| D | b | 1912 | CCGGACACAGGCGCGCGCTTGGGCTCAACCGCGCCCTCGAGTCAAGATGCACTCGGCG | 1853 |
| O | y | 661 | TCAGGCTGAGTCAAGGCTGAGCTTCTACTCTGGGAGCAACAAGGACCTCTCGAGGCT | 720 |
| D | b | 1852 | TCAGGCTGAGTCAAGGCTGAGCTTCTACTCTGGGAGCAACAAGGACCTCTCGAGGCT | 1793 |
| O | y | 721 | ACCAACAACCCGAGCAGCAAGACATCCACCAACTTCCAGCCACCACTCTGATCT | 780 |
| D | b | 1792 | ACCAACAACCCGAGCAGCAAGACATCCATTTCAATTCACGACCAACACTCTGATCT | 1733 |
| O | y | 781 | CCTACCAACCTTGGCAGCGCATACACAAGAATGATGCGATGACCTGACCATAGACG | 840 |
| D | b | 1732 | CCTACCAACCTTGGCAGCGCATACACAAGAATGATGCGATGACCTGACCATAGACG | 1673 |
| O | y | 841 | GTAACCTCCCTCAACCTCTCCATACAGACACTTCTCCGAGTTGCTPACTGGGCTCT | 900 |
| D | b | 1672 | GTAACCTCTCTCAACCTCTCCATACAGACACTTCTCCGAGTTGCTPACTGGGCTCTCT | 1613 |
| O | y | 901 | TTCTTTTTCCTGCTCTTTCAATTTCAAACTCCAGTTAATTCCTCTGSAAGATCCC | 960 |
| D | b | 1612 | TTCTTTTTCCTGCTCTTTCAATTTCAAACTCCAGTTAATTCCTCTGSAAGATCCC | 1553 |
| O | y | 961 | AGCAGCAGTACTACCAAGAGCTGAGAGAGACATTTCTGAAATGTTTTTGGAGATTTAT | 1020 |
| D | b | 1552 | AGCAGCAGTACTACCAAGAGCTGAGAGAGACATTTCTGAAATGTTTTTGGAGATTTAT | 1493 |
| O | y | 1021 | AAACAAGGGGGTTTTCTGGGCGCTCTCCATATATTAAGTCAAGCGCAGGATCTGGATGTA | 1080 |
| D | b | 1492 | AAACAAGGGGGTTTTCTGGGCGCTCTCCATATATTAAGTCAAGCGCAGGATCTGGATGTA | 1493 |
| O | y | 1081 | CAATTGACTCTGGCGCTTCCAGAAAGTCAATCAATGTCCACGACGTGGAGACAACAGTTC | 1140 |
| D | b | 1432 | CAATTGACTCTGGCGCTTCCAGAAAGTCAATCAATGTCCACGACGTGGAGACAACAGTTC | 1373 |
| O | y | 1141 | AATCAGTATPAAACGAAAGCGGCTCTCGATTTAACTGAGATCTCAGACGTCAAGCTG | 1200 |
| D | b | 1372 | AATCAGTATPAAACGAAAGCGGCTCTCGATTTAACTGAGATCTCAGACGTCAAGCTG | 1313 |
| O | y | 1201 | AGTACATGTCCAATTCCTTTCTGCCACAGTCTGGGGCTGGGGTGCAGGCTGGGCAATC | 1260 |
| D | b | 1312 | AGTACATGTCCAATTCCTTTCTGCCACAGTCTGGGGCTGGGGTGCAGGCTGGGCAATC | 1253 |
| O | y | 1261 | GCGCTCTGATGCTGATCTGATGTTCTGATGTGCGCTGGGCAATGTCTATCTCAATGTGCTTG | 1320 |
| D | b | 1252 | GCGCTCTGATGCTGATCTGATGTTCTGATGTGCGCTGGGCAATGTCTATCTCAATGTGCTTG | 1193 |
| O | y | 1321 | GCTGCTGTCACATGCGCGCCGAAAGAACTPACGGGAGCGTGAACATCTTCCAGCGCGGAGT | 1380 |
| D | b | 1192 | GCTGCTGTCACATGCGCGCCGAAAGAACTPACGGGAGCGTGAACATCTTCCAGCGCGGAGT | 1133 |
| O | y | 1381 | ACCTTACATCTATGAGCGAGTACCCOACTTACACACATCCATGGGGGCTATGTGCCCTT | 1440 |
| D | b | 1132 | ACCTTACATCTATGAGCGAGTACCCOACTTACACACATCCATGGGGGCTATGTGCCCTT | 1073 |
| O | y | 1441 | AGCAGTACCGATCTGATGCCCTTATGAAAGATTTCTGAGGTAATGATGCGACAGCTCT | 1500 |
| D | b | 1072 | AGCAGTACCGATCTGATGCCCTTATGAAAGATTTCTGAGGTAATGATGCGACAGCTCT | 1013 |
| O | y | 1501 | TTCTTACAAACCGAGAGTGGAGCGCACTTCTGCAACTTGTATGGGGGCA | 1550 |
| D | b | 1012 | TTCTTACAAACCGAGAGTGGAGCGCACTTCTGCAACTTGTATGGGGGCA | 963 |

RESULT 15
US-09-726-172-2117/c

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 23:14:01 ; Search time 74 Seconds
(without alignments)
16.206 Million cell updates/sec

```
Title: US-09-658-621B-26
Perfect score: 44
Sequence: 1 ISEMFLOIY 9
```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 130866

| | | | | |
|---------|----|-----|---------|---|
| Minimum | DB | seq | length: | 0 |
| Maximum | DB | seq | length: | 9 |

Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_GeneSeq_101002:.*

- 1: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1960.DAT.*
- 2: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1961.DAT.*
- 3: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1962.DAT.*
- 4: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1963.DAT.*
- 5: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1964.DAT.*
- 6: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1965.DAT.*
- 7: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1966.DAT.*
- 8: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1967.DAT.*
- 9: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1968.DAT.*
- 10: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1969.DAT.*
- 11: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneSeq/geneSeqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 44 | 100.0 | 9 | 22 | AAU00563 | Human MUC1 polypep |
| 2 | 22 | 50.0 | 9 | 19 | AAW59466 | HIV-1 Rev protein |
| 3 | 22 | 50.0 | 9 | 19 | AAW59469 | HIV-1 Rev protein |
| 4 | 22 | 50.0 | 9 | 21 | AAV58724 | Paraneoplastic cer |
| 5 | 22 | 50.0 | 9 | 22 | AA895813 | p53 epitope A2 sup |
| 6 | 22 | 50.0 | 9 | 22 | AA895885 | p53 epitope A2 sup |
| 7 | 22 | 50.0 | 9 | 22 | AA893706 | p53 epitope HLA-A2 |
| 8 | 22 | 50.0 | 9 | 23 | AA905955 | p53 epitope HLA-A2 |
| 9 | 21 | 47.7 | 8 | 16 | AAK73578 | Immunogenic peptid |
| 10 | 21 | 47.7 | 8 | 21 | AA08585 | Human ISH receptor |
| 11 | 21 | 47.7 | 8 | 21 | AA08585 | Peptide identified |

| | | | | | | |
|----|----|------|---|----|-----------|----------------------|
| 11 | 21 | 47.7 | 9 | 14 | AAR44265 | Residues 129-137 of |
| 12 | 21 | 47.7 | 9 | 15 | AAR44282 | Residues 129-137 of |
| 13 | 21 | 47.7 | 9 | 15 | AAV18357 | PAF-derived HLA-B1 |
| 14 | 21 | 47.7 | 9 | 15 | AAR59189 | Peptide fragment (|
| 15 | 21 | 47.7 | 9 | 17 | AAAR97530 | Antigenic peptide, |
| 16 | 21 | 47.7 | 9 | 20 | AAV53361 | p53 epitope (aa 12 |
| 17 | 21 | 47.7 | 9 | 20 | AAV45938 | Immunogenic peptide, |
| 18 | 21 | 47.7 | 9 | 20 | AAV40184 | Amino acid sequenc |
| 19 | 21 | 47.7 | 9 | 20 | AAV6700 | HLA-A2 binding p53 |
| 20 | 21 | 47.7 | 9 | 21 | AAAB3703 | Cytotoxic T lympho |
| 21 | 21 | 47.7 | 9 | 21 | AAAB21704 | Cytotoxic T lympho |
| 22 | 21 | 47.7 | 9 | 21 | AAAB67152 | Cancer testis tumo |
| 23 | 21 | 47.7 | 9 | 22 | AAAG89424 | p53 DR supermotif |
| 24 | 21 | 47.7 | 9 | 22 | AAAG89532 | p53 epitope A2 sup |
| 25 | 21 | 47.7 | 9 | 22 | AAAG89584 | p53 epitope A2 sup |
| 26 | 21 | 47.7 | 9 | 22 | AAAG89705 | p53 epitope HLA-A* |
| 27 | 21 | 47.7 | 9 | 22 | AAAB69924 | Human NY-ESO-1 HLA- |
| 28 | 21 | 47.7 | 9 | 22 | AAAB76012 | Tumour associated |
| 29 | 21 | 47.7 | 9 | 22 | AAAR13643 | HIV protease inhibi |
| 30 | 20 | 45.5 | 5 | 12 | AAAR13646 | HIV protease inhibi |
| 31 | 20 | 45.5 | 7 | 22 | AAAM46601 | H11 binding site c |
| 32 | 20 | 45.5 | 7 | 22 | AAAM62244 | H11 binding site c |
| 33 | 20 | 45.5 | 7 | 22 | AAAM62249 | H11 binding site c |
| 34 | 20 | 45.5 | 7 | 22 | AAAM63378 | H11 binding site c |
| 35 | 20 | 45.5 | 8 | 15 | AAAR70990 | Bovine ER receptor |
| 36 | 20 | 45.5 | 8 | 15 | AAV38031 | Hepatitis B virus- |
| 37 | 20 | 45.5 | 9 | 15 | AAV38032 | Hepatitis B virus- |
| 38 | 20 | 45.5 | 9 | 20 | AAV43601 | Immunogenic peptid |
| 39 | 20 | 45.5 | 9 | 20 | AAV03004 | Fragment of human |
| 40 | 20 | 45.5 | 9 | 22 | AAE05206 | Human HLA-A1 bindi |
| 41 | 20 | 45.5 | 9 | 23 | AAU94465 | Human novel protea |
| 42 | 20 | 45.5 | 9 | 23 | AAU94852 | Human novel protea |
| 43 | 19 | 43.2 | 5 | 12 | AAAR13644 | HIV protease inhibi |
| 44 | 19 | 43.2 | 5 | 12 | AAAR13645 | HIV protease inhibi |
| 45 | 19 | 43.2 | 6 | 11 | AAAR08214 | HIV protease and 1 |

ALIGNMENTS

RESULT 1
AAU00563
ID AAU00563 standard; Peptide; 9 AA

AC AAU00563

DT 12-SEP-2001 (first entry)

DE Human MUC1 polypeptide derivative #24.

KW Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-1

KW cancer gene therapy; diagnosis; treatment; inflammatory disorder
KW organ transplant rejection; graft versus host disease.

KW organ transplant rejection; graft versus host disease..

Os Homo sapiens

PN W0200118035-A2

PD 15-MAR-2001.

PF 07-SEP-2000; 2000WO-EP08761

| | | |
|----|--------------|--------------|
| PR | 08-SEP-1999; | 99GB-0021242 |
| PR | 10-SEP-1999; | 99FB-0402237 |

PR 03-MAR-2000; 2000US-0187215

PA (IRGB) IRANSGENE SA.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD

Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM

XX

XX

DR WPI: 2001-235187/24.
DR N-PSDB; AAS00609.

PT New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer

PS Claim 2; Page 74; 81pp; English.

CC The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumours.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.

CC Sequence 9 AA;

Query Match 100.0%; Score 44; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLOIY 9
DB 1 ISEMFLOIY 9

RESULT 2

AAMS59466
ID AAMS59466 standard; peptide; 9 AA.

XX AAMS59466;

XX 28-AUG-1998 (first entry)

XX HIV-1 Rev protein HLA-A1 peptide motif #1 from patient L658.

XX Human immunodeficiency virus; HIV; AIDS; cytotoxic T cell; lymphocyte;
KW acquired immunodeficiency syndrome; Rev protein; epitope; immunisation;
KW detection; prognosis; diagnosis; treatment; prevention.

XX Human immunodeficiency virus type 1.

XX WO9817309-A1.

XX 30-APR-1998.

XX 17-OCT-1997; 97WO-IB01402.

XX 18-OCT-1996; 96US-0733789.

XX (UYRO-) UNIV ROTTERDAM ERASMUS.

XX Osterhaus ADME, Van Baalen CA;

XX WPI; 1998-261189/23.

PT Use of Rev or Tat proteins of immunodeficiency virus - for
PT developing products for diagnosis, prognosis, prevention or
PT treatment of infection, especially HIV infection

PS Disclosure; Page 19; 38pp; English.

CC AAMS59466-W59480 are fragments of the human immunodeficiency virus Rev
CC protein which are used in a novel method to detect at least 1 cytotoxic
CC protein epitope of the virus in order to prevent the disease. The
CC specification also describes a method of immunising a host against
CC disease caused by infection by an immunodeficiency virus, which
CC comprises stimulating in a host a specific cytotoxic T-cell response
CC which is specific for the Rev and/or Tat proteins of the immunodeficiency
CC virus. A method determining favourable prognosis in an HIV positive
CC subject which comprises detecting the presence of a cytotoxic T-cell
CC response to Rev and/or Tat HIV protein as an indication of the favourable
CC the diagnosis, monitoring, prevention and treatment of immunodeficiency
CC disease mediated by an immunodeficiency virus, especially HIV.

CC Sequence 9 AA;

Query Match 50.0%; Score 22; DB 19; Length 9;
Best Local Similarity 55.6%; Pred. No. 7.8e+05;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ISEMFLOIY 9
DB 1 ISEMFLOIY 9

RESULT 3

AAMS59469
ID AAMS59469 standard; peptide; 9 AA.

XX AAMS59469;

XX 28-AUG-1998 (first entry)

XX HIV-1 Rev protein HLA-A1 peptide motif #1 from patient P424.

XX Human immunodeficiency virus; HIV; AIDS; cytotoxic T cell; lymphocyte;
KW acquired immunodeficiency syndrome; Rev protein; epitope; immunisation;
KW detection; prognosis; diagnosis; treatment; prevention.

XX Human immunodeficiency virus type 1.

XX WO9817309-A1.

XX 30-APR-1998.

XX 17-OCT-1997; 97WO-IB01402.

XX 18-OCT-1996; 96US-0733789.

XX (UYRO-) UNIV ROTTERDAM ERASMUS.

XX Osterhaus ADME, Van Baalen CA;

XX WPI; 1998-261189/23.

PT Use of Rev or Tat proteins of immunodeficiency virus - for
PT developing products for diagnosis, prognosis, prevention or
PT treatment of infection, especially HIV infection

PS Disclosure; Page 19; 38pp; English.

CC AAMS59466-W59480 are fragments of the human immunodeficiency virus Rev
CC protein which are used in a novel method to detect at least 1 cytotoxic
CC protein epitope of the virus in order to prevent the disease. The
CC specification also describes a method of immunising a host against
CC disease caused by infection by an immunodeficiency virus, which
CC comprises stimulating in a host a specific cytotoxic T-cell response
CC which is specific for the Rev and/or Tat proteins of the immunodeficiency
CC virus. A method determining favourable prognosis in an HIV positive
CC subject which comprises detecting the presence of a cytotoxic T-cell
CC response to Rev and/or Tat HIV protein as an indication of the favourable

CC prognosis is also described. The compositions and methods can be used for
 CC the diagnosis, monitoring, prevention and treatment of immunodeficiency
 CC disease mediated by an immunodeficiency virus, especially HIV.

XX
 SQ Sequence 9 AA;

Query Match 50.0%; Score 22; DB 19; Length 9;
 Best Local Similarity 55.6%; Pred. No. 7.8e+05;
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ISEMFLQY 9
 : |||
 Db 1 ISERLISTY 9

RESULT 4
 AAY58724
 ID AAY58724 standard; Peptide; 9 AA.
 AC AAY58724;

XX 25-APR-2000 (first entry)

DE Paraneoplastic cerebellar degeneration antigen cdr2 peptide cdr2-2.

XX Paraneoplastic cerebellar degeneration; PCD; cdr2-2; Yoz; human;

KW cellular immunity; immune privileged antigen;

KM cytotoxic T lymphocyte; tumour; autoimmune disease; therapy;

XX antitumour; immunomodulator.

OS Homo sapiens.

XX WO20000825-A2.

PD 06-JAN-2000.

XX 30-JUN-1999; 99WO-US14827.

XX 30-JUN-1998; 98US-0107878.

PR 29-JUN-1999; 99US-0107978.

XX (UYRQ) UNIV ROCKEFELLER.

PA Darnell RB, Albert ML, Bhardwaj N;

PI WPI; 2000-170934/15.

XX DR detecting, quantifying and modulating cellular immune response to

PT immune-privileged antigens, for diagnosis and treatment of, e.g. tumors

PT and paraneoplastic diseases -

XX Claim 40; Page 73; 83pp; English.

XX The present sequence is that of peptide cdr2-2 (or Yoz), 1 of 8

XX claimed peptides (see AAY58723-30), termed cdr2-1 to cdr2-8 (Yoz-Yoz8),

XX derived from cdr2. The paraneoplastic cerebellar degeneration (PCD)

XX antigen has been identified as cdr2, a protein expressed in neuronal

XX cells and in gynaecological tumours. Investigation into the

XX processing of the cdr2 antigen for presentation by antigen-presenting

XX cells to T-lymphocytes has led to the identification of polypeptide

XX fragments of cdr2 which are targets for naturally-occurring

XX cytotoxic T-lymphocytes (CTLs) in PCD patients. These peptides are

XX believed to be those presented by dendritic cells in the development

XX of cellular immunity. The invention relates to methods and agents

XX for the detection and modulation of cellular immunity to immune

XX privileged antigens. Antigen-specific CTLs are present in subjects

XX with PND which is now explained in terms of expression, in remote

XX tissues, of the same antigen as expressed by a tumour. Loss of the

XX immune-privileged status of the antigen, e.g. as the result of an

XX effective antitumour response, may cause an autoimmune disease.

XX Agents that decrease the level of antigen-specific T cells are used

XX to treat PND, and antigen-expressing non-tumour cells can be

XX protected against antigen-specific T cells by treating them with an

CC agent that reduces their sensitivity to T cells. Increasing the
 CC number of antigen-specific CTLs is used for treatment of
 CC antigen-expressing neoplasms.

XX
 SQ Sequence 9 AA;

Query Match 50.0%; Score 22; DB 21; Length 9;
 Best Local Similarity 66.7%; Pred. No. 7.8e+05;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFL 6
 : |||
 Db 3 ISEMFL 8

RESULT 5
 AAG89533
 ID AAG89533 standard; Peptide; 9 AA.
 AC AAG89533;

XX 11-SEP-2001 (first entry)

DE p53 epitope A2 supermotif analogue peptide #17.

XX Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;

KW cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;

XX vaccine; epitope; cytostatic.

OS Homo sapiens.

XX WO200141788-A1.

PD 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33629.

XX 10-DEC-1999; 99US-0458297.

PR (EPI-M-) EPIMMUNE INC.

PA Pike J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

PI Keogh E;

XX WPI; 2001-381493/40.

XX DR EpiTope-based vaccines comprising p53 epitope having a specified

PT sequences, useful for treating and preventing cancer, the epitopic

PT peptides is useful as diagnostic agents and for evaluating immune

XX response -

XX Example 4; Page 118; 138pp; English.

XX The present invention describes isolated prepared p53 epitopes (I). Also

XX described are: (1) a cytotoxic T lymphocyte (CTL) that is cultured

XX in vivo and binds to a complex of (I); (2) a peptide (II) comprising (1)

XX and a second epitope and has less than 50 contiguous amino acids; (3) a

XX vaccine composition comprising (II), a unit dose of a peptide with less

XX than 50 contiguous amino acids with 100% identity to the native peptide

XX sequence of p53, and a pharmaceutical excipient; (4) an isolated nucleic

XX acid encoding (I); and (5) an isolated nucleic acid encoding (II). (I)

XX has cytostatic activity and can be used in vaccines. The vaccine

XX composition is useful for treating or preventing cancer. (I) and (II)

XX are useful as diagnostic agents and for evaluating immune responses.

XX Unlike conventional epitopes, immunosuppressive epitopes that may be

XX present in whole antigens can be avoided with the use of the vaccine

XX composition of (II). The ability to combine selected epitopes and

XX further, to modify the composition of the epitopes enhances the

XX immunogenicity. The possible pathological side effects caused by

XX infectious agents or whole protein antigens, which might have their own

XX intrinsic biological activity, are eliminated. AAG89363 to AAG89747

XX represent amino acid sequences used in the exemplification of the

CC present invention.

XX Sequence 9 AA;

Query Match 50.0%; Score 22; DB 22; Length 9;
Best Local Similarity 37.5%; Pred. No. 7.8e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
:::|::|:
DB 2 LNKMFQCV 9

RESULT 6

AAAG89585 ID AAG89585 standard; Peptide; 9 AA.

XX AAG89585;

DT 11-SEP-2001 (first entry)

DE p53 epitope A2 supermotif crossbinding peptide #20.

XX Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;
XX cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;
XX vaccine; epitope; cytostatic.

XX Homo sapiens.

OS Synthetic.

XX WO200141788-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33629.

XX 10-DEC-1999; 99US-0458297.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;

XX WPI; 2001-381493/40.

PT Epitope-based vaccines comprising p53 epitope having a specified
PT sequences, useful for treating and preventing cancer, the epitopic
PT peptides is useful as diagnostic agents and for evaluating immune
PT response

PS Example 4; Page 120; 138pp; English.

CC The present invention describes isolated prepared p53 epitopes (I). Also
CC described are: (1) a cytotoxic T lymphocyte (CTL) that is cultured
CC in vivo and binds to a complex of (I); (2) a peptide (II) comprising (I)
CC and a second epitope and has less than 50 contiguous amino acids; (3) a
CC vaccine composition comprising (II), a unit dose of a peptide with less
CC than 50 contiguous amino acids with 100% identity to the native peptide
CC sequence of p53; and a pharmaceutical excipient; (4) an isolated nucleic
CC acid encoding (I); and (5) an isolated nucleic acid encoding (II). (I)
CC has cytostatic activity and can be used in vaccines. The vaccine
CC composition is useful for treating or preventing cancer. (I) and (II)
CC are useful as diagnostic agents and for evaluating immune responses.
CC Unlike conventional epitopes, immunosuppressive epitopes that may be
CC present in whole antigens can be avoided with the use of the vaccine
CC composition of (I). The ability to combine selected epitopes and
CC further, to modify the composition of the epitopes enhances the
CC immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigens, which might have their own
CC intrinsic biological activity, are eliminated. AAG89363 to AAG89747
CC represent amino acid sequences used in the exemplification of the
CC present invention.

SQ Sequence 9 AA;

Query Match 50.0%; Score 22; DB 22; Length 9;
Best Local Similarity 37.5%; Pred. No. 7.8e+05;
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
:::|::|:
DB 2 LNKMFQCV 9

RESULT 7

AAAG89706 ID AAG89706 standard; Peptide; 9 AA.

XX AAG89706;

DT 11-SEP-2001 (first entry)

DE p53 epitope HLA-A2 supermotif peptide #5.

XX Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;
XX cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;
XX vaccine; epitope; cytostatic.

XX Homo sapiens.

OS Synthetic.

XX WO200141788-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33629.

XX 10-DEC-1999; 99US-0458297.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;

XX WPI; 2001-381493/40.

PT Epitope-based vaccines comprising p53 epitope having a specified
PT sequences, useful for treating and preventing cancer, the epitopic
PT peptides is useful as diagnostic agents and for evaluating immune
PT response

PS Claim 1; Page 128; 138pp; English.

CC The present invention describes isolated prepared p53 epitopes (I). Also
CC described are: (1) a cytotoxic T lymphocyte (CTL) that is cultured
CC in vivo and binds to a complex of (I); (2) a peptide (II) comprising (I)
CC and a second epitope and has less than 50 contiguous amino acids; (3) a
CC vaccine composition comprising (II), a unit dose of a peptide with less
CC than 50 contiguous amino acids with 100% identity to the native peptide
CC sequence of p53; and a pharmaceutical excipient; (4) an isolated nucleic
CC acid encoding (I); and (5) an isolated nucleic acid encoding (II). (I)
CC has cytostatic activity and can be used in vaccines. The vaccine
CC composition is useful for treating or preventing cancer. (I) and (II)
CC are useful as diagnostic agents and for evaluating immune responses.
CC Unlike conventional epitopes, immunosuppressive epitopes that may be
CC present in whole antigens can be avoided with the use of the vaccine
CC composition of (I). The ability to combine selected epitopes and
CC further, to modify the composition of the epitopes enhances the
CC immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigens, which might have their own
CC intrinsic biological activity, are eliminated. AAG89363 to AAG89747
CC represent amino acid sequences used in the exemplification of the
CC present invention.

Query Match 50.0%; Score 22; DB 22; Length 9;
 Best Local Similarity 37.5%; Pred. No. 7.8e+05;
 Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
 DB 2 LNMFCOV 9

RESULT 8

AAU95955

ID AAU95955 standard; Peptide; 9 AA.

AC AAU95955;

DT 02-JUL-2002 (first entry)

DE Immunogenic peptide with (HLA)-A2.1 binding site #168.

XX HLA-A2.1 binding peptide; cytostatic; virucide; anti-HIV; hepatotropic;
 KW human immunodeficiency virus; anti-inflammatory; antibacterial; vaccine;
 KW procoagulant; immunosuppressant; immunogenic peptide; T cell activation;
 KW human leucocyte antigen binding site; cytotoxic T cell response;
 KW viral infection; hepatitis; Epstein-Barr virus; papilloma virus;
 KW human immunodeficiency virus; HIV; Kaposi sarcoma; Lassa fever virus;
 KW cytomegalovirus; tumour; prostate cancer; renal carcinoma; lymphoma;
 KW prostate-specific antigen; p53; carcino-embryonal antigen;
 KW melanoma antigen; Mycobacterium tuberculosis; protozoa;
 KW trypanosome surface antigen; condyloma acuminatum.

XX Unidentified.

OS WO200220616-A1.

PN 14-MAR-2002.

PD 01-SEP-2000; 2000MO-US24102.

PF 01-SEP-2000; 2000MO-US24102.

PR (EPIM-) EPIMUNE INC.

PA Grey HM, Sette A, Sidney J, Southwood S;

PI WPI; 2002-351766/38.

PT Immunogenic peptide with human leucocyte antigen-A2.1 binding site,

PS useful for treating e.g. viral infection or tumours -

XX Claim 1; Page 29; 35pp; English.

CC The invention describes a composition comprising an immunogenic peptide
 CC having a human leucocyte antigen (HLA)-A2.1 binding site. The peptides
 CC bind specifically to HLA-A2.1, to cause T cell activation and thus a
 CC cytotoxic T cell response. The peptides and the nucleic acids that
 CC encode them, are used, in vivo or ex vivo, for treatment of viral
 CC infections (hepatitis B or C; Epstein-Barr; human immune deficiency;
 CC Kaposi sarcoma; human papilloma; Lassa fever or cytomegalovirus);
 CC tumours including prostate cancer, renal carcinoma and lymphoma (where
 CC directed to prostate-specific antigen, p53, carcino-embryonal antigen,
 CC Her2/neu or melanoma antigens); infection by Mycobacterium tuberculosis
 CC or protozoa (directed to trypanosome surface antigen); and condyloma
 CC acuminatum. The peptides are suitable for use in peptide-based
 CC vaccines. This sequence represents an immunogenic peptide with the
 CC human leucocyte antigen (HLA)-A2.1 binding site, described in the
 CC invention.

XX Sequence 9 AA;

Query Match 50.0%; Score 22; DB 23; Length 9;
 Best Local Similarity 37.5%; Pred. No. 7.8e+05;
 Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 8
 DB 2 LNMFCOV 9

RESULT 9

AAR73578

ID AAR73578 standard; Peptide; 8 AA.

AC AAR73578;

DT 19-DEC-1995 (first entry)

DE Human TSH receptor (residues 755-762).

XX thyrioid stimulating hormone receptor; TSH; human; Homo sapiens;
 KW antibody; affinity; detection.

OS Synthetic.

PN JP07089991-A.

PD 04-APR-1995.

PF 28-SEP-1993; 93JP-0240853.

PR 28-SEP-1993; 93JP-0240853.

PA (MITP) MITSUBISHI PETROCHEMICAL CO LTD.

DR WPI; 1995-167251/22.

XX Novel polypeptide(s) having affinity for the human TSH receptor
 PT antibody - used in detection of the TSH antibody.

XX Example 1; Page 51; 54pp; Japanese.

CC Peptides with affinity to human TSH (thyrioid stimulating hormone)
 CC receptor antibody are used for detection of the antibody. (See also
 CC AAR73201-592).

SQ Sequence 8 AA;

Query Match 47.7%; Score 21; DB 16; Length 8;
 Best Local Similarity 57.1%; Pred. No. 7.8e+05;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLOI 7
 DB 1 ISEMFLOI 7

RESULT 10

ID AAB08585 standard; Peptide; 8 AA.

AC AAB08585;

DT 20-DEC-2000 (first entry)

DE Peptide identified from an origin of DNA polymerase epsilon.

XX Precursor peptide; polypeptide hormone; peptide identification.

XX Unidentified.

OS Unidentified.

PN Key

FT Modified-site

PN WO200050636-A1.

```
XX 31-AUG-2000.
PD
XX
XX 24-FEB-2000; 2000WO-FR00460.
PF
XX 25-FEB-1999; 99US-0257525.
PR
XX (SCRC ) SCRAS SOC CONSEILS RECH & APPL SCT.
PA (CNRS ) CNRS CENT NAT RECH SCT.
XX Camara Ferrer YJA, Thuriere C, Martinez J, Borge G, Goze C,
XX WPI; 2000-572101/53.
DR
XX
XX Identifying peptide with selected function, useful particularly for
PT C-amidated hormones, by screening database for combination of nucleic
PT acid and amino acid sequences
XX
XX Disclosure; Page 21; 40pp; French.
XX
XX The specification describes a method for identifying a peptide having
XX a particular function. The method comprises preparing a database of
XX polynucleotides and polypeptides of unknown functions, screening the
XX database for a combination of nucleotides or amino acids indicative of
XX the peptide with a particular function, and identifying polynucleotides
XX and proteins which contain the peptide. The method is used to identify
XX precursor peptides with an amidated C-terminus, especially polypeptide
XX hormones, for studying physiologically active substances. The present
XX sequence represents a peptide which was identified using the method of
XX the invention.
XX
XX Sequence 8 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 21; Length 8;
XX Best Local Similarity 80.0%; Pred. No. 7.8e+05;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISEMF 5
Db 1 ISELF 5
XX
XX RESULT 11
XX AAR44265
XX ID AAR44265 standard; peptide; 9 AA.
XX
XX AAR44265;
XX 09-JUN-1994 (first entry)
XX
XX Residues 129-137 of human p53.
XX
XX Human; p53; protein; Class I; allele; diagnosis; assay; detection;
XX p53 protein-specific; T-cell; antibody; cancer; MHC; HLA-A2.1.
XX
XX Homo sapiens.
XX
XX WO9324525-A.
XX
XX 09-DEC-1993.
XX
XX 18-MAY-1993; 93WO-NL00102.
XX
XX 26-MAY-1992; 92EP-0201510.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (SEED-) SEED CAPITAL INVESTMENTS (SCI) BV.
XX
XX Kast WM, Melief CJM;
XX WPI; 1993-405730/50.
XX
XX Peptide derived from p53 protein - used to treat and diagnose
PT
```

```
PT diseases involving over-expression of p53 e.g. human cancers
XX
XX Claim 3; Page 37; 46pp; English.
XX
XX The sequences given in AAR4426-80 are peptides derived from human p53
XX protein which have the ability to bind to a human Class I molecule.
XX These peptides may be used in a diagnostic test or assay to detect
XX human p53 protein-specific T-cells or antibodies. They may also be
XX used in the treatment of diseases such as human cancers showing p53
XX protein overexpression. These peptides can bind human MHC Class I
XX allele HLA-A2.1.
XX
XX Sequence 9 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 14; Length 9;
XX Best Local Similarity 37.5%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOI 8
Db 2 INKMFCDL 9
XX
XX RESULT 12
XX AAR44282
XX ID AAR44282 standard; peptide; 9 AA.
XX
XX AAR44282;
XX
XX 09-JUN-1994 (first entry)
XX
XX Residues 129-137 of human mp53.
XX
XX Human; p53; protein; Class I; allele; diagnosis; assay; detection;
XX p53 protein-specific; T-cell; antibody; cancer; MHC; HLA-A2.1.
XX
XX Homo sapiens.
XX
XX WO9324525-A.
XX
XX 09-DEC-1993.
XX
XX 18-MAY-1993; 93WO-NL00102.
XX
XX 26-MAY-1992; 92EP-0201510.
XX
XX (UYLE-) RIJKSUNIV LEIDEN.
XX (SEED-) SEED CAPITAL INVESTMENTS (SCI) BV.
XX
XX Kast WM, Melief CJM;
XX WPI; 1993-405730/50.
XX
XX Peptide derived from p53 protein - used to treat and diagnose
XX diseases involving over-expression of P53 e.g. human cancers
XX
XX Claim 3; Page 38; 46pp; English.
XX
XX The sequences given in AAR44281-94 are peptides derived from human mp53
XX protein which have the ability to bind to a human Class I molecule.
XX These peptides may be used in a diagnostic test or assay to detect
XX human p53 protein-specific T-cells or antibodies. They may also be
XX used in the treatment of diseases such as human cancers showing p53
XX protein overexpression. These peptides can bind human MHC Class I
XX allele HLA-A2.1.
XX
XX Sequence 9 AA;
SQ
XX
XX Query Match 47.7%; Score 21; DB 14; Length 9;
XX Best Local Similarity 37.5%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISEMFLOI 8
```

Db 2 LNKMFWYQL 9

RESULT 13
ID AAY38357
XX AAY38357 standard; Peptide; 9 AA.

AC AAY38357;

DT 29-SEP-1999 (first entry)

DE PAP-derived HLA-binding peptide.

XX Immunogen: HLA; human leukocyte antigen; binding motif; antiviral;
KW MHC; major histocompatibility complex; viral infection; anticancer;
KM prostate cancer; lymphoma; hepatitis; AIDS; diagnostic; diagnosis.

OS Homo sapiens.

XX MO9403205-A1.

XX 17-FEB-1994.

XX 06-AUG-1993; 93WO-US07421.

XX 05-MAR-1993; 93US-0027746.

XX 07-AUG-1992; 92US-0926666.

XX (CYTE-) CYTEL CORP.

XX Celis E, Grey HM, Kubo RT, Sette A;

XX WPI; 1994-065403/08.

XX Peptide which specifically binds selected MHC allele - used to
PT induce an immune response for treatment or prevention of viral
XX infection or cancer, or for diagnosis

XX Disclosure; Page 114; 150pp; English.

XX The sequence is a specific example of a group of new immunogenic
CC peptides having an HLA-A3.2, HLA-A1, HLA-A11 or HLA-A24.1 binding
CC motif. For example, the peptides having an HLA-A3.2 binding motif
CC each have 9-10 residues and contain, from the N-terminus to the
CC C-terminus, (a) a first conserved residue selected from L, M, I,
CC V, S, A, T, F, C, G, D and E and (b) a second conserved residue of
CC K, R, Y, H or F, where the first and second conserved residues are
CC separated by 6-7 residues. The peptides are capable of binding
CC selected MHC molecules and inducing an immune response. They can be
CC used to treat and/or prevent viral infection and cancer, e.g. prostate
CC cancer, lymphoma, hepatitis or AIDS. They can also be used to produce
CC antibodies for use as diagnostic or therapeutic agents. The peptides
CC can also be used as diagnostic agents.

XX Sequence 9 AA;

XX Query Match 47.7%; Score 21; DB 15; Length 9;
XX Best Local Similarity 42.9%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMFLQIY 9
XX | | | | |
XX 3 EYFVENY 9

RESULT 14

ID AAR59189 standard; peptide; 9 AA.

AC AAR59189;

DT 03-MAY-1995 (first entry)

XX Peptide fragment (1.0277) of p53 binds HLA-A2.1.

XX antigen; epitope; immunogenic target protein; PSA; HBVC; HBVB; EBV;
KW HIV1; core antigen; surface antigen; pharmaceutical composition;
KW in vivo; ex vivo; therapeutic; diagnostic; MHC class I molecule;
KM major histocompatibility complex; HLA-A2.1; 9mer; 10mer; anchor;
XX human leukocyte antigen.

OS Homo sapiens.

XX MO9420127-A.

XX 15-SEP-1994.

XX 04-MAR-1994; 94WO-US02353.

XX 05-MAR-1993; 93US-0027746.

XX 04-JUN-1993; 93US-0073205.

XX 29-NOV-1993; 93US-0159184.

XX (CYTE-) CYTEL CORP.

XX Grey HM, Kaat WM, Sette A, Sidney J;

XX WPI; 1994-302678/37.

XX Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT for treatment or prophylaxis of cancer, virus infection or
XX autoimmune diseases.

XX Example 5; Page 103; 138pp; English.

XX AAR59104-264 are immunogenic 9mer peptides that contain a HLA-A2.1
CC binding motif. These peptides bind HLA-A2.1 and have a binding affinity
CC of at least 1% as compared to a reference peptide (AAR71293). AAR59189
CC has an IC50 of 0.0013 and the sequence occurs at position 129 in the p53
CC protein. The peptides of the invention can induce cytotoxic T lymphocytes
CC which can react with target cells. They can be used for the treatment or
CC prophylaxis of cancer, eg. prostate cancer or lymphoma, etc.

XX Sequence 9 AA;

XX Query Match 47.7%; Score 21; DB 15; Length 9;
XX Best Local Similarity 37.5%; Pred. No. 7.8e+05;
XX Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISEMFLQI 8
XX | | | | |
XX 2 LNKMFWCQL 9

RESULT 15

ID AAR97530 standard; peptide; 9 AA.

AC AAR97530;

DT 12-FEB-1997 (first entry)

DE Antigenic peptide, corresp. to p53 aa 129-137.

XX p53; Her-2; Neu; aa; amino acid; CTL; cytotoxic T lymphocyte; target;
KW malignant cell; antigenic; vaccine; immunisation; activation.

XX Homo sapiens.

XX MO9618409-A1.

XX 20-JUN-1996.

XX 14-DEC-1995; 95WO-US16415.

PR 14-DEC-1994; 94US-0355558.

XX (SCRI) SCRIpps RES INST.

XX Sherman LA;

XX WPI; 1996-300385/30.

XX In vivo activation of tumour-specific cytotoxic T lymphocytes - by
PT contacting with polypeptide(s) derived from human p53 or Her-2/Neu
PT proteins

XX Example 4; Page 94; 158pp; English.

XX AA97517-R97544 are antigenic p53-derived peptides synthesised to be
CC used in an assay for identifying peptides capable of activating
CC cytotoxic T lymphocytes (CTLs) which specifically target malignant
CC cells. CTL-activating peptides can be used in a vaccine for
CC protecting against tumour cell formation. CTLs activated by the
CC peptides will lyse tumour cells displaying specific peptides.
CC Antibodies against CTL-activating peptides are useful for the
CC identification of other similar compounds which may be useful for
CC treating cancer or virally-infected cells, or for diagnosis. The
CC peptide and vaccines produced provide immunity to a high percentage
CC of different ethnic groups, i.e. those with different HLA alleles.

XX SQ Sequence 9 AA;

Query Match 47.7%; Score 21; DB 17; Length 9;

Best Local Similarity 37.5%; Pred. No. 7.8e+05;

Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 1SEMFLQI 8

DB 2 LNRMECOL 9

Search completed: May 1, 2003, 23:22:57
Job time : 75 secs

GenCore version 5.1.4.D5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: May 8, 2003, 14:56:22 ; Search time 73 Seconds
(without alignments)
6604.053 Million cell updates/sec

Title: US-09-658-621B-1
Perfect score: 1572
Sequence: 1 gattccctgagctgctgaa.....tcgccctgagctgagcgtg 1572

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|---------|---------------------|--------------------|
| 1 | 1196.6 | 76.1 | 6192 | 2 US-08-479-537A-1 | Sequence 1, Appli |
| 2 | 1196.6 | 76.1 | 6192 | 4 US-09-083-116-1 | Sequence 1, Appli |
| 3 | 1196.6 | 76.1 | 6192 | 4 US-09-134-916A-1 | Sequence 1, Appli |
| 4 | 625.2 | 39.8 | 6449 | 2 US-08-479-537A-4 | Sequence 4, Appli |
| 5 | 625.2 | 39.8 | 6449 | 4 US-09-083-116-4 | Sequence 4, Appli |
| 6 | 625.2 | 39.8 | 6449 | 4 US-09-134-916A-4 | Sequence 4, Appli |
| 7 | 77.2 | 5.0 | 320 | 4 US-09-165-264-13 | Sequence 13, Appli |
| 8 | 77.2 | 4.9 | 320 | 4 US-09-165-264-12 | Sequence 12, Appli |
| 9 | 77.2 | 4.9 | 320 | 4 US-09-165-264-7 | Sequence 7, Appli |
| 10 | 76.6 | 4.9 | 320 | 4 US-09-165-264-14 | Sequence 14, Appli |
| 11 | 76.6 | 4.8 | 319 | 4 US-09-165-264-8 | Sequence 8, Appli |
| 12 | 75.8 | 4.8 | 320 | 4 US-09-128-155-16 | Sequence 16, Appli |
| 13 | 73.8 | 4.7 | 152331 | 3 US-09-103-840A-2 | Sequence 2, Appli |
| 14 | 70.4 | 4.5 | 4403765 | 4 US-09-103-840A-2 | Sequence 11, Appli |
| 15 | 69.8 | 4.4 | 12001 | 1 US-08-458-568A-11 | Sequence 11, Appli |
| 16 | 67.8 | 4.3 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appli |
| 17 | 67.7 | 4.3 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appli |
| 18 | 65.8 | 4.2 | 1505 | 1 US-07-915-246-1 | Sequence 1, Appli |
| 19 | 62.6 | 4.0 | 494 | 4 US-09-056-556-176 | Sequence 176, App |
| 20 | 62.6 | 4.0 | 494 | 4 US-09-072-596-171 | Sequence 171, App |
| 21 | 62.6 | 4.0 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, Appli |
| 22 | 61.2 | 3.9 | 985 | 4 US-09-056-556-182 | Sequence 182, App |
| 23 | 61.2 | 3.9 | 985 | 4 US-09-072-596-177 | Sequence 177, App |
| 24 | 61 | 3.9 | 1926 | 4 US-09-249-585A-4 | Sequence 4, Appli |
| 25 | 61 | 3.9 | 1931 | 2 US-09-130-114-2 | Sequence 2, Appli |
| 26 | 60.2 | 3.8 | 530 | 3 US-08-758-662-4 | Sequence 4, Appli |
| 27 | 60.2 | 3.8 | 4524 | 2 US-08-845-998-7 | Sequence 7, Appli |

| | | | | | |
|------|------|-----|---------|---------------------|--------------------|
| C 28 | 60.2 | 3.8 | 4524 | 3 US-09-206-537-7 | Sequence 7, Appli |
| C 29 | 60.2 | 3.8 | 4524 | 4 US-09-430-854-7 | Sequence 7, Appli |
| C 30 | 59 | 3.8 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, Appli |
| C 31 | 58.4 | 3.7 | 6152 | 2 US-08-479-537A-1 | Sequence 1, Appli |
| C 32 | 58.4 | 3.7 | 6152 | 4 US-09-083-116-1 | Sequence 1, Appli |
| C 33 | 58.4 | 3.7 | 6192 | 4 US-09-134-916A-1 | Sequence 4, Appli |
| C 34 | 58.4 | 3.7 | 6449 | 2 US-08-479-537A-4 | Sequence 4, Appli |
| C 35 | 58.4 | 3.7 | 6449 | 4 US-09-083-116-4 | Sequence 4, Appli |
| C 36 | 58.4 | 3.7 | 6449 | 4 US-09-134-916A-4 | Sequence 4, Appli |
| C 37 | 58.4 | 3.7 | 33529 | 4 US-09-144-085-3 | Sequence 3, Appli |
| C 38 | 58 | 3.7 | 2367 | 4 US-09-056-556-201 | Sequence 201, App |
| C 39 | 58 | 3.7 | 2367 | 4 US-09-072-596-196 | Sequence 196, App |
| C 40 | 57.4 | 3.7 | 2186 | 2 US-08-878-546-9 | Sequence 9, Appli |
| C 41 | 57.4 | 3.7 | 6530 | 2 US-08-146-930-1 | Sequence 1, Appli |
| C 42 | 57.4 | 3.7 | 6530 | 3 US-08-458-240-1 | Sequence 1, Appli |
| C 43 | 57.4 | 3.7 | 6530 | 5 PCT-US93-0393-1 | Sequence 1, Appli |
| C 44 | 57.4 | 3.7 | 13987 | 2 US-08-804-227C-13 | Sequence 13, Appli |
| C 45 | 57.4 | 3.7 | 44377 | 2 US-08-804-227C-7 | Sequence 7, Appli |

ALIGNMENTS

RESULT 1
US-08-479-537A-1
Sequence 1, Application US/08479537A
Patent No. 5861381/
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVIENT, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs

```

? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: sig peptide
? LOCATION: 58..120
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 439..5239
? OTHER INFORMATION: /note= "The nucleotides spanning
? OTHER INFORMATION: 439-5239 constitute a repeated region wherein the repeat is 6
? OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
? OTHER INFORMATION: The number of such repeats varies from 1 to 80."
? FEATURE:
? NAME/KEY: mat peptide
? LOCATION: 121..6166
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 457
? OTHER INFORMATION: /note= "Nucleotide 457 is XI = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
? OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 487
? OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
? OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
? OTHER INFORMATION: or ACG, and Asn = AAT or AAC."
? FEATURE:
? NAME/KEY: repeat_region
? LOCATION: 496
? OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
? OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
? OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
? US-08-479-537A-1

Query Match 76.1%; Score 1196.6; DB 2; Length 6192;
Best Local Similarity 92.2%; Pred. No. 7e-267;
Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 246 CCACAGCCCCGGGTTCAGGCTCTCCACCACTCAGAGGACAGGATGTCACTCTGGCCCCGGC 305
DB 4866 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 4925
QY 306 CACGAAACGAGCTTCAGTTCAGTGTGACCTGGGAGACAGATGTCACTCTGGCTCCAGT 365
DB 4926 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 4985
QY 366 CACAGGCCAGGCTTGAGCTTCCACACCGCCGACGATGTCACTCTGAGCCCCGGG 425
DB 4986 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 5045
QY 426 CACAGGCCAGGCTTGAGCTTCCACACCGCCGACGAGTGTCACTCTGGCCCCGGG 485
DB 5046 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 5105
QY 486 CACAGGCCAGGCTTGAGCTTCCACACCGCCGACGAGTGTCACTCTGGCCCCGGG 545
DB 5106 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 5165
QY 546 CACAGGCCAGGCTTGAGCTTCCACACCGCCGACGAGTGTCACTCTGGCCCCGGG 605
DB 5166 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 5225
QY 606 CACAGGCCAGGCTTGAGCTTCCACACCGCCGACGAGTGTCACTCTGGCCCCGGG 665
DB 5226 CANNAGGCCGANNCCGGGCTCTCACCGCCCCCANNNGCCACGAGTGTCACTCTGGCCCCGGG 5285
QY 666 CTCGTGATAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTCTGAGAGGCTACAC 725
DB 5286 CTCGTGATAGGCTCAGCTTCTACTCTGTGTGACAAAGGACCTCTGAGAGGCTACAC 5345
```

```

QY 726 AACCCAGCCAGCAAGACATCTCACCCAGCATTCGCCAGCCACCTCTGATCTCTAC 785
DB 5346 AACCCAGCCAGCAAGACATCTCACCCAGCATTCGCCAGCCACCTCTGATCTCTAC 5405
QY 786 CACCTTGGCAGCATATACACACAGACTGATGCGAGTAGCACTACCATATGACGGTACC 845
DB 5406 CACCTTGGCAGCATATACACACAGACTGATGCGAGTAGCACTACCATATGACGGTACC 5465
QY 846 TCTCTGACCTCTCCATTCACAGACATCTTCCCAATTCTCTAGAGGAGTCTTTCTT 905
DB 5466 TCTCTGACCTCTCCATTCACAGACATCTTCCCAATTCTCTAGAGGAGTCTTTCTT 5525
QY 906 TTTCTGCTCTTTTCACATTTTCAACCTTCAGATTAAATTCCTCTGAGAAATCCAGAC 965
DB 5526 TTTCTGCTCTTTTCACATTTTCAACCTTCAGATTAAATTCCTCTGAGAAATCCAGAC 5585
QY 966 CGACTACTACAGAGGCTGACAGAGACATTTCTGAAATGTTTTGGAGATTATTAACA 1025
DB 5586 CGACTACTACAGAGGCTGACAGAGACATTTCTGAAATGTTTTGGAGATTATTAACA 5645
QY 1026 AGGGGGTTTTCTGGGCTCTCTCAATTTAATTAGTTGAGCCAGAGATCTGTGGTGTACAT 1085
DB 5646 AGGGGGTTTTCTGGGCTCTCTCAATTTAATTAGTTGAGCCAGAGATCTGTGGTGTACAT 5705
QY 1086 GACTCTGGGCTTCCAGAAAGGTACATCATATGTCCACGACGTGAGACACAGTTCAATCA 1145
DB 5706 GACTCTGGGCTTCCAGAAAGGTACATCATATGTCCACGACGTGAGACACAGTTCAATCA 5765
QY 1146 GTATTAACAGAAAGGCTCTGATTTAATCTGACGATCTCAGACGTGAGTGTCA 1205
DB 5766 GTATTAACAGAAAGGCTCTGATTTAATCTGACGATCTCAGACGTGAGTGTCA 5825
QY 1206 TGGGCAATTTCTCTCTGCGGAGTCTGGGCTGGGGTGGGAGCTGGGGGACATGGGCT 1265
DB 5826 TGGGCAATTTCTCTCTGCGGAGTCTGGGCTGGGGTGGGAGCTGGGGGACATGGGCT 5885
QY 1266 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1325
DB 5886 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5945
QY 1326 CTGTGAGTGGCCGCAAGAACTACAGGGGAGCTGACATCTTTTCAGCCCGGAGTACTCA 1385
DB 5946 CTGTGAGTGGCCGCAAGAACTACAGGGGAGCTGACATCTTTTCAGCCCGGAGTACTCA 6005
QY 1386 CCAATCTTAGAGAGATACCCCACTACACACCCAGAGGCTAGTATGTGCCCCCTAGAGAG 1445
DB 6006 CCAATCTTAGAGAGATACCCCACTACACACCCAGAGGCTAGTATGTGCCCCCTAGAGAG 6065
QY 1446 TACCATGTGAGCCCTTAGAGAAAGTTTCTGAGGTATATGTGTGACAGGCTCTCTTA 1505
DB 6066 TACCATGTGAGCCCTTAGAGAAAGTTTCTGAGGTATATGTGTGACAGGCTCTCTTA 6125
QY 1506 CACAAACCCAGAGGTGAGGACATTTCTGCAATTTGTATGGGAGACGCTGCGCTGAGC 1565
DB 6126 CACAAACCCAGAGGTGAGGACATTTCTGCAATTTGTATGGGAGACGCTGCGCTGAGC 6185
QY 1566 TGAAGTGG 1572
DB 6186 TGAAGTGG 6192

RESULT 2
US-09-083-116-1
; Sequence 1, Application US/0908116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mera
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
```


CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DONALD, SWECKER & MATTHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 58..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 439..5239
OTHER INFORMATION: /note= "The nucleotides spanning 439-5239 constitute a repeated region wherein the repeat is 6
OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
OTHER INFORMATION: The number of such repeats varies from 1 to 80."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 121..6166
FEATURE:
NAME/KEY: repeat_region
LOCATION: 457
OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 487
OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: repeat_region
LOCATION: 496
OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

US-09-083-116-1

Query Match 76.1%; Score 1196.6; DB 4; Length 6192;
Best Local Similarity 92.2%; Pred. No. 7e-267;
Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| QY | 246 | CCACAGCCCGGCTTCAGGCTCTCTCCACCACTCAGAGAGAGATGTACTCTGGCCCCCGGC | 305 |
| DB | 4866 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 4925 |
| QY | 306 | CACGGAACCAAGCTTCAGGTTGAGCTGCCACCTGGGGACAGAGATGATCACTCGATCCAGT | 365 |
| DB | 4926 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 4985 |
| QY | 366 | CACCAAGCCAGCCCTGGGCTCCACCACTCCCGCAGCCCAAGATGTACTCTGAGCCCCGGA | 425 |
| DB | 4986 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 5045 |
| QY | 426 | CAACAGCCCGGCTCCACCGCCCCCAGGCTCCACGAGTCACTCGGCCCGGA | 485 |
| DB | 5046 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 5105 |
| QY | 486 | CACCAAGCCCGGCTCCACCGCCCCCAGGCTCCACGAGTCACTCGGCCCGGA | 545 |
| DB | 5106 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 5165 |
| QY | 546 | CACCAAGCCCGGCTCCACCGCCCCCAGGCTCCACGAGTCACTCGGCCCGGA | 605 |
| DB | 5166 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 5225 |
| QY | 606 | CACCAAGCCCGGCTCCACCGCCCCCAGGCTCCACGAGTCACTCGGCCCGGA | 665 |
| DB | 5226 | CNNNAGGCCGNNNCCGGGCTCCACCGCCCCCANNNGCCACGAGTCACTCGGCCCGGA | 5285 |
| QY | 666 | CTCTGATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTCTGCGAGGCTACAC | 725 |
| DB | 5286 | CTCTGATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTCTGCGAGGCTACAC | 5345 |
| QY | 726 | AACCCAGCCAGCAAGACCTCCAGGCTTCCAGGCGACCACTCTGATACCTCTAC | 785 |
| DB | 5346 | AACCCAGCCAGCAAGACCTCCAGGCTTCCAGGCGACCACTCTGATACCTCTAC | 5405 |
| QY | 786 | CACCCAGCCAGCAAGACCTCCAGGCTTCCAGGCGACCACTCTGATACCTCTAC | 845 |
| DB | 5406 | CACCCAGCCAGCAAGACCTCCAGGCTTCCAGGCGACCACTCTGATACCTCTAC | 5465 |
| QY | 846 | TCCTCTACCTCTCCAAATCAAGCACTTCTCCAGTTGTTACTGGGCTCTTTT | 905 |
| DB | 5466 | TCCTCTACCTCTCCAAATCAAGCACTTCTCCAGTTGTTACTGGGCTCTTTT | 5525 |
| QY | 906 | TTTCGCTTTTTCATTTCAATTTCAAACTCCAGTTAATCTCTCTGGAAGATCCAGCAC | 965 |
| DB | 5526 | TTTCGCTTTTTCATTTCAATTTCAAACTCCAGTTAATCTCTCTGGAAGATCCAGCAC | 5585 |
| QY | 966 | CGACTACTACCAAGCTGCAAGAGACATTTTGAATGTTTTCGATTTTAAACA | 1025 |
| DB | 5586 | CGACTACTACCAAGCTGCAAGAGACATTTTGAATGTTTTCGATTTTAAACA | 5645 |
| QY | 1026 | AGGGGGTTTTTCGGGCTCTCCAAATTTAAGTTCAAGGCCAGGATCTGTGTATCAATT | 1085 |
| DB | 5646 | AGGGGGTTTTTCGGGCTCTCCAAATTTAAGTTCAAGGCCAGGATCTGTGTATCAATT | 5705 |
| QY | 1086 | GACTCTGGCTTCCAGAAAGTATCAATGATTCACGACGTTGAGACAGGTTCAATCA | 1145 |
| DB | 5706 | GACTCTGGCTTCCAGAAAGTATCAATGATTCACGACGTTGAGACAGGTTCAATCA | 5765 |
| QY | 1146 | GTTATTAAGGAAAGAGGCTCTCGATATACTGACGATCTGACGCTCAAGCTGAGTCA | 1205 |
| DB | 5766 | GTTATTAAGGAAAGAGGCTCTCGATATACTGACGATCTGACGCTCAAGCTGAGTCA | 5825 |
| QY | 1206 | TGTGCAATTTCTTTTCTGCTGCCAGCTGGGGCTGGGGGCGAGGCTGGGCACTGGGCT | 1265 |
| DB | 5826 | TGTGCAATTTCTTTTCTGCTGCCAGCTGGGGCTGGGGGCGAGGCTGGGCACTGGGCT | 5885 |

Thu May 8 18:53:07 2003

us-09-658-621b-1.rni

Page 6

```

; NAME/KEY: repeat_region
; LOCATION: 496
; OTHER INFORMATION: /note="Nucleotide 496 is X2 = NNN
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG, and Ala = GCT, GCC, GCA, or GCG."
;
; US-08-479-537A-4

```

| | | | | |
|-----------------------|-------|--------------------|------|-------------|
| Query Match | 39.8% | Score 625.2 | DB 2 | Length 6449 |
| Best Local Similarity | 92.0% | Pred. No. 3.8e-135 | | |

[illegible]

RESULT 5
US-09-083-116-4
Sequence 4, Application US/09083116
Patent No.:6203795
GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENEY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVENT, Maya
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

```

1 ADDRESSSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
2 STREET: P.O. Box 1404
3 CITY: Alexandria
4 STATE: Virginia
5 COUNTRY: United States
6 ZIP: 22313-1404
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patentin Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/09/083,116
14 FILING DATE:
15 CLASSIFICATION:
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: 08/479,537
18 FILING DATE:
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: WO PCT/FR91/00835
21 FILING DATE: 23-OCT-1991
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/039,320
24 FILING DATE: 04-APR-1993
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/403,576
27 FILING DATE: 14-MAR-1995
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Teskin, Robin L.
30 REGISTRATION NUMBER: 35,030
31 REFERENCE/DOCKET NUMBER: 017753-025
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (703) 836-6620
34 TELEFAX: (703) 836-2021
35 INFORMATION FOR SEQ ID NO: 4:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 6449 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41 MOLECULE TYPE: DNA (genomic)
42 FEATURE:
43 NAME/KEY: sig_peptide
44 LOCATION: 58..120
45 FEATURE:
46 NAME/KEY: repeat_region
47 LOCATION: 439..5239
48 OTHER INFORMATION: /note= "The nucleotides spanning
49 439-5239 constitute a repeated region wherein the repeat is 6
50 OTHER INFORMATION: nucleotides and encodes 20 amino acids, 17 of which are fixed
51 OTHER INFORMATION: The number of such repeats varies from 1 to 80."
52 FEATURE:
53 NAME/KEY: mat_peptide
54 LOCATION: 121..5661
55 FEATURE:
56 NAME/KEY: repeat_region
57 LOCATION: 457
58 OTHER INFORMATION: /note= "Nucleotide 457 is X1 = NNN
59 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
60 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
61 NAME/KEY: repeat_region
62 LOCATION: 487
63 OTHER INFORMATION: /note= "Nucleotide 487 is Y = NNN
64 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
65 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
66 FEATURE:
67 NAME/KEY: repeat_region
68 LOCATION: 496
69 OTHER INFORMATION: /note= "Nucleotide 496 is X2 = NNN
70 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
71 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866

```



```
QY 1 GAATTCCTGCTGCTGTAATCTGTTCTGCCCCCTTCCCACTTTCACCACTTTCACCACTG 60
Db 1 GAATTCCTGCTGCTGTAATCTGTTCTGCCCCCTTCCCACTTTCACCACTTTCACCACTG 60
QY 61 AACACGGGAGCAGCAGTCTCCCTTCTTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTT 120
Db 61 AACACGGGAGCAGCAGTCTCCCTTCTTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTT 120
QY 121 ACAAGTTCTGCTCATGCAAGCTCTACCCAGAGTGAAGAAAGAGACTTGGCTACCCAG 180
Db 121 ACAAGTTCTGCTCATGCAAGCTCTACCCAGAGTGAAGAAAGAGACTTGGCTACCCAG 180
QY 181 AGAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 181 AGAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 241 TCCAGCCACAGCCCGGCTTCAAGCTCTCTCCACCACTCAAGGAGACAGATGTCACTGAGCC 300
Db 241 TCCAGCCACAGCCCGGCTTCAAGCTCTCTCCACCACTCAAGGAGACAGATGTCACTGAGCC 300
QY 301 CCGGCGACAGGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTCACTGAGCTC 360
Db 301 CCGGCGACAGGAAACAGCTTCAAGTTCAAGTCACTGAGGAGACAGATGTCACTGAGCTC 360
QY 361 CCAAGTCAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 420
Db 361 CCAAGTCAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 420
QY 421 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 480
Db 421 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 480
QY 481 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 540
Db 481 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 540
QY 541 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 600
Db 541 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 600
QY 601 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 660
Db 601 CCGGACAAACAGGCTCAGCTTGGGCTTCCACCACTCGGCGACGCTCAAGATGTCACTGAGCC 660
QY 661 TCAGGCTCTGCTCAGGCTCAGCTTCTTAC 689
Db 661 TCAGGCTCTGCTCAGGCTCAGCTTCTTAC 689
```

```
RESULT 7
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinyagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13
```

```
Query/ Match 5.0%; Score 78; DB 4; Length 320;
Best Local Similarity 53.2%; Pred. No. 2,4e-09;
Matches 165; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
```

```
QY 353 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 412
Db 313 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 254
QY 413 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 472
Db 253 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 194
QY 473 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 532
Db 193 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 134
QY 533 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 592
Db 133 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 74
QY 593 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 652
Db 73 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 14
QY 653 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 710
Db 13 CTTGAGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCA 4
```

```
RESULT 8
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinyagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12
```

```
Query Match 4.9%; Score 77.2; DB 4; Length 318;
Best Local Similarity 53.3%; Pred. No. 3.7e-09;
Matches 163; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
```

```
QY 357 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 416
Db 315 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 256
QY 417 AGCCCGGAGCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 476
Db 255 AGCCCGGAGCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 196
QY 477 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 536
Db 195 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 136
QY 537 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 596
Db 135 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 76
QY 597 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 656
Db 75 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 16
QY 657 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 716
Db 15 GGTCCAGTCAACAGGACGCTGAGCTTCCACCACTCGGCGACGCTTCAAGTGTCACTC 4
```


GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2003, 14:07:27 ; Search time 4354 Seconds

(without alignments)
10507.490 Million cell updates/sec

Title: US-09-658-621B-1

Perfect score: 1572

Sequence: 1 gattcccttgctgctgaa.....tcgccccttgagctgagtg 1572

Scoring table: IDENTITY NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_scs.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_scs.*

27: em_sy.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rdd.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|--------------|---------------------|
| 1 | 1572 | 100.0 | 1572 | 6 AX093798 | AX093798 Sequence |
| 2 | 1481.2 | 94.2 | 1721 | 6 AX335860 | AX335860 Sequence |
| 3 | 1481.2 | 94.2 | 1721 | 6 AX440427 | AX440427 Sequence |
| 4 | 1481.2 | 94.2 | 1721 | 9 HSTRYMA | X52229 Human mRNA |
| 5 | 1412.2 | 89.8 | 1804 | 6 AX335367 | AX335367 Sequence |
| 6 | 1412.2 | 89.8 | 1804 | 6 HDMWUCAB | J05581 Human polyom |
| 7 | 1328.4 | 84.5 | 1795 | 6 GIBWUC1A | L41589 Hylobates 1 |
| 8 | 1313 | 83.5 | 1452 | 6 A32135 | A32135 H.sapiens m |
| 9 | 1268 | 80.7 | 1834 | 12 AF423030 | AF423030 Synthetic |
| 10 | 1196.6 | 76.1 | 6192 | 6 AR030811 | AR030811 Sequence |
| 11 | 1196.6 | 76.1 | 6192 | 6 AR142537 | AR142537 Sequence |
| 12 | 1190.2 | 75.7 | 4139 | 6 AX334899 | AX334899 Sequence |
| 13 | 1190.2 | 75.7 | 4139 | 6 AX335372 | AX335372 Sequence |
| 14 | 1190.2 | 75.7 | 4139 | 6 AX336712 | AX336712 Sequence |
| 15 | 1190.2 | 75.7 | 4139 | 6 AX409474 | AX409474 Sequence |
| 16 | 1190.2 | 75.7 | 4139 | 6 AX440481 | AX440481 Sequence |
| 17 | 1190.2 | 75.7 | 4139 | 6 HUMPANMU | J05582 Human pancr |
| 18 | 1190.2 | 75.7 | 4139 | 6 HUMEPISIA2 | M34089 Human episi |
| 19 | 998 | 63.5 | 1296 | 9 HUMEPISIB2 | E08764 cDNA encodi |
| 20 | 998 | 63.5 | 1296 | 9 HUMEPISIB2 | X52228 Human mRNA |
| 21 | 993.4 | 63.2 | 1320 | 6 E08764 | AX406624 Sequence |
| 22 | 919.4 | 58.5 | 2238 | 9 HSBETA | M61170 Human polyom |
| 23 | 750 | 47.7 | 8181 | 6 HUMPEM | AL771399 Homo sapi |
| 24 | 746.4 | 47.5 | 1709 | 9 A32137 | AL771399 Homo sapi |
| 25 | 745.2 | 47.4 | 178229 | 2 AL713399 | L41624 Hylobates 1 |
| 26 | 745.2 | 47.4 | 178229 | 2 AL713399 | U60261 Human MUC-1 |
| 27 | 742 | 47.2 | 3343 | 9 HUMETMAGA | AR030812 Sequence |
| 28 | 694 | 44.1 | 5660 | 9 GIBWUC1A01 | AR142538 Sequence |
| 29 | 670.2 | 42.6 | 859 | 9 HSB0261 | U60261 Human MUC-1 |
| 30 | 627 | 39.9 | 2068 | 4 BTA400824 | AR030812 Sequence |
| 31 | 625.2 | 39.8 | 6449 | 6 AR030812 | AR142538 Sequence |
| 32 | 625.2 | 39.8 | 6449 | 6 AR142538 | U60261 Human MUC-1 |
| 33 | 619.4 | 39.4 | 805 | 9 HSB0260 | X80761 H.sapiens M |
| 34 | 607.4 | 38.6 | 768 | 9 HSMUC1 | AF125525 Homo sapi |
| 35 | 605.8 | 38.5 | 759 | 9 AF125525 | U85787 Oryctolagus |
| 36 | 600.4 | 38.2 | 2119 | 4 OC885787 | BC005441 Mus muscu |
| 37 | 586.2 | 37.3 | 2242 | 10 BC005441 | M84683 Mus muscu |
| 38 | 584 | 37.2 | 2166 | 10 MOSWUC1A | U36918 Mesocricetu |
| 39 | 556.8 | 35.4 | 2370 | 10 MAU36918 | AF176947 Macaca mu |
| 40 | 531 | 33.8 | 2751 | 9 AF176947 | U60259 Human MUC-1 |
| 41 | 495.4 | 31.5 | 730 | 9 HSB0259 | AX440341 Sequence |
| 42 | 494.4 | 31.5 | 497 | 6 AX440341 | M32739 Human episi |
| 43 | 482.6 | 30.7 | 572 | 9 HUMEPISIB1 | M31823 Human brea |
| 44 | 475.6 | 30.3 | 572 | 9 HUMEPISIB1 | AX347259 Sequence |
| 45 | 459 | 29.2 | 5660 | 6 AX347259 | |

ALIGNMENTS

RESULT 1

AX093798

LOCUS AX093798 1572 bp DNA linear PAT 30-MAR-2001

DEFINITION Sequence 1 from Patent WO0118035.

ACCESSION AX093798

VERSION AX093798.1 GI:13510057

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1572)

AUTHORS Taylor-Papadimitriou,J., Heukamp,L.C., Offringa,R., Melief,C.J., Acres,B. and Thomas,M.

TITLE Muc-1 derived peptides

REFERENCE 1
 AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G., Horvitz, S., Soppet, D.R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature gene sets
 JOURNAL Patent: WO 0194629-A 6369 13-DEC-2001;
 AVALON Pharmaceuticals (US)
 FEATURES Location/Qualifiers
 source 1..1721
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 339 a 634 c 395 g 353 t
 ORIGIN

Query Match 94.2% Score 1481.2; DB 6; Length 1721;
 Best Local Similarity 95.5%; Pred. No. 1.8e-260;
 Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 1 GAATTCCTGGCTGTTGAATCTGTTCCCTCCCAACCATTTCCACCAACCATG 60
 DB 1 GAATTCCTGGCTGTTGAATCTGTTCCCTCCCAACCATTTCCACCAACCATG 60
 QY 61 ACACCGGGACCCAGTCTCTCTTCTCTGCTGCTCTCTCAAGTGTTCAGTTGT 120
 DB 61 ACACCGGGACCCAGTCTCTCTTCTCTGCTGCTCTCTCAAGTGTTCAGTTGT 120
 QY 121 ACAGGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGACTTCGGCTAC 180
 DB 121 ACAGGTTCTGTCATGCAAGCTCTACCCAGGTGGAGAAAGAGACTTCGGCTAC 180
 QY 181 AGAAGTTCAGTCCCAAGCTCTACGAGAAAGATGCTGTAGATGACAGAGGTA 240
 DB 181 AGAAGTTCAGTCCCAAGCTCTACGAGAAAGATGCTGTAGATGACAGAGGTA 240
 QY 241 TCCAGCCACAGCCCGGCTTCAAGGCTCTCCACCACTCAGGAGACAGATGTAC 300
 DB 241 TCCAGCCACAGCCCGGCTTCAAGGCTCTCCACCACTCAGGAGACAGATGTAC 300
 QY 301 CCGGCGACCGGAACAGCTTCAAGTTCAGCTGCCACTGGGGAACAGATGTAC 360
 DB 301 CCGGCGACCGGAACAGCTTCAAGTTCAGCTGCCACTGGGGAACAGATGTAC 360
 QY 361 CCAATCACCAGGCTCAGCTCTGAGCAACCCGCGACAGCCAGATGTACCTGAG 420
 DB 361 CCAATCACCAGGCTCAGCTCTGAGCAACCCGCGACAGCCAGATGTACCTGAG 420
 QY 421 CCGGACAAACAGCTCAGCTCCGCGGCTCCACCGCCCGGCGACAGATGTAC 480
 DB 421 CCGGACAAACAGCTCAGCTCCGCGGCTCCACCGCCCGGCGACAGATGTAC 480
 QY 481 CCGGACAAACAGCTCAGCTCCGCGGCTCCACCGCCCGGCGACAGATGTAC 513
 DB 481 CCGGACAAACAGCTCAGCTCCGCGGCTCCACCGCCCGGCGACAGATGTAC 513
 QY 514 -----CCGCGCGCCACAGTGTACCTTGAGCC 540
 DB 541 CCGGACCAAGGCTCGCCCGGCGCTCCACCGCGCGCCAGGATGTACCTTGAG 600
 QY 541 CCGGACCAAGGCTCGCCCGGCGCTCCACCGCGCGCCAGGATGTACCTTGAG 600
 DB 601 CCGGACCAAGGCTCGCCCGGCGCTCCACCGCGCGCCAGGATGTACCTTGAG 660
 QY 601 CCGGACCAAGGCTCGCCCGGCGCTCCACCGCGCGCCAGGATGTACCTTGAG 660
 DB 661 CCGGACCAAGGCTCGCCCGGCGCTCCACCGCGCGCCAGGATGTACCTTGAG 720
 QY 661 TGAAGCTTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTGCGAG 720
 DB 661 TGAAGCTTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTGCGAG 720
 QY 721 TGAAGCTTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTGCGAG 780
 DB 721 TGAAGCTTGCATCAGGCTCAGCTTCTACTGTGTCACAAAGGACCTTGCGAG 780
 QY 721 ACCACAAACCCAGGACGAGAGAGACCTCCACCGAGATTCCAGCCACTCTGAT 780

DB 781 ACCACAAACCCAGGACGAGAGAGACCTCCACCGAGATTCCAGCCACTCTGAT 840
 QY 781 CTTACCACTCTTCCAGGACATAGACCAAGACTGATGCGAGTACACTCACAAT 840
 DB 841 CTTACCACTCTTCCAGGACATAGACCAAGACTGATGCGAGTACACTCACAAT 900
 QY 841 GATACCTCTCAGCTCCTCCATATACAGACACTTCTCCGAGTGTCTACTGGG 900
 DB 901 GATACCTCTCAGCTCCTCCATATACAGACACTTCTCCGAGTGTCTACTGGG 960
 QY 901 TTTCTTTTCTGCTCTTTTCAATTCAAACTCCAGTTTAATTCCTCTGGAAGAT 960
 DB 961 TTTCTTTTCTGCTCTTTTCAATTCAAACTCCAGTTTAATTCCTCTGGAAGAT 1020
 QY 961 AGACCGACTCTACCAAGAGCTGACAGAGACATTTGAAATGTTTGGCAGATT 1020
 DB 1021 AGACCGACTCTACCAAGAGCTGACAGAGACATTTGAAATGTTTGGCAGATT 1080
 QY 1021 AAAAAGGGGGTTTTCTGGGCTCTCCAAATATTAAATTCAAGCCAGATCTG 1080
 DB 1081 AAAAAGGGGGTTTTCTGGGCTCTCCAAATATTAAATTCAAGCCAGATCTG 1140
 QY 1081 CAATTGACTCTGGCTTCCGAGAGGTACATCAATGTCACAGCGTGGAGACAG 1140
 DB 1141 CAATTGACTCTGGCTTCCGAGAGGTACATCAATGTCACAGCGTGGAGACAG 1200
 QY 1141 AATCAGTATTAACGGAAGAGCTCTCGATATAAAGCTGACATCTCAGACG 1200
 DB 1201 AATCAGTATTAACGGAAGAGCTCTCGATATAAAGCTGACATCTCAGACG 1260
 QY 1201 AGTCATGTGCAATTCCTTTCTCTGCGCAGCTGGGGCTGGGGCTGCGAGC 1260
 DB 1261 AGTCATGTGCAATTCCTTTCTCTGCGCAGCTGGGGCTGGGGCTGCGAGC 1320
 QY 1261 GCGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 DB 1321 GCGCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 QY 1321 GCTGTCTGTAGTGCAGCGGAAAGAACTACGCGGAGCTGAGACATCTTTC 1380
 DB 1381 GCTGTCTGTAGTGCAGCGGAAAGAACTACGCGGAGCTGAGACATCTTTC 1440
 QY 1381 ACTTACCATCTTATGAGAGAGTACCCCACTTACCAACCCATGAGGCTG 1440
 DB 1441 ACTTACCATCTTATGAGAGAGTACCCCACTTACCAACCCATGAGGCTG 1500
 QY 1441 AGCAGTACCGATGTAGGCTCTATGAGAAAGTTTCTGCAAGTAAATGTG 1500
 DB 1501 AGCAGTACCGATGTAGGCTCTATGAGAAAGTTTCTGCAAGTAAATGTG 1560
 QY 1501 TCTTACCAAAACCCAGACATGAGACCACTTCTGCAACTTGTGAGGGGCA 1560
 DB 1561 TCTTACCAAAACCCAGACATGAGACCACTTCTGCAACTTGTGAGGGGCA 1620
 QY 1561 TGAAGTGAATG 1572
 DB 1621 TGAAGTGAATG 1632

RESULT 3
 AX440427 1721 bp DNA linear PAT 28-JUN-2002
 LOCUS AX440427
 DEFINITION Sequence 280 from Patent WO0190154.
 ACCESSION AX440427
 VERSION AX440427.1 GI:21665237
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1
 Xu, J., Mitcham, J.L., Harlocker, S.L., Dillon, D.C., Secret, H.,

Lodes,M.J., Algate,P.A., Fling,S.P., Mannion,J., Benson,D.R. and Carter,D.
Compositions and methods for the therapy and diagnosis of ovarian cancer

JOURNAL Patent: WO 0190154-A 280 29-NOV-2001;
CORIXA CORPORATION (US)

FEATURES Location/Qualifiers

Source 1..1721

/db_xref="taxon:9606"

BASE COUNT 339 a 634 c 395 g 353 t

ORIGIN

Query Match 94.2%; Score 1481.2; DB 6; Length 1721;

Best Local Similarity 95.5%; Pred. No. 1.8e-260;

Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

```

QY 1 GAATTCCTGGCTGCTTGAATCTGTCTGCCCCCTCCCAACCATTTCAACACACATG 60
DB 1 GAATTCCTGGCTGCTTGAATCTGTCTGCCCCCTCCCAACCATTTCAACACACATG 60
QY 61 AACCGGGGCAACCGCTGCTTCTTCCGCTGCTGCTTCAAGTCTTCAAGTCTT 120
DB 61 AACCGGGGCAACCGCTGCTTCTTCTTCTGCTGCTTCTCAAGTCTTCAAGTCTT 120
QY 121 AACAGTTCTGTCTCAATGCAAGCTTCAACCCAGGTGAGAAAAGAGACTTGGCTACG 180
DB 121 AACAGTTCTGTCTCAATGCAAGCTTCAACCCAGGTGAGAAAAGAGACTTGGCTACG 180
QY 181 AAGAGTTCAAGTCCAGCTCTACTAGAGAAATGCTGAGATATGACAGACGCTACTC 240
DB 181 AAGAGTTCAAGTCCAGCTCTACTAGAGAAATGCTGAGATATGACAGACGCTACTC 240
QY 241 TCCAGCCACAGCCCGGTTTCAAGCTCTTCAACCATTCAGGGAAGATGTCATCTGGCC 300
DB 241 TCCAGCCACAGCCCGGTTTCAAGCTCTTCAACCATTCAGGGAAGATGTCATCTGGCC 300
QY 301 CCGGCGACGGAACAGCTTCAAGTTCAGCTGCACTTGGGAGACAGATGTCATCTGGCT 360
DB 301 CCGGCGACGGAACAGCTTCAAGTTCAGCTGCACTTGGGAGACAGATGTCATCTGGCT 360
QY 361 CCAATGACACAGGCTGAGCTTCAACCATTCAGGGAAGATGTCATCTGGCT 420
DB 361 CCAATGACACAGGCTGAGCTTCAACCATTCAGGGAAGATGTCATCTGGCT 420
QY 421 CCGGCAACAAGCCAGCCCGGCTTCAACCGCCCGCGCCCAAGGTGTCATCTGGCC 480
DB 421 CCGGCAACAAGCCAGCCCGGCTTCAACCGCCCGCGCCCGCGCCCAAGGTGTCATCTGGCC 480
QY 481 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 513
DB 481 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 513
QY 514 -----CCGCGCGCCCAAGGTGTCATCTGGCC 540
DB 514 -----CCGCGCGCCCAAGGTGTCATCTGGCC 540
QY 541 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 600
DB 541 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 600
QY 601 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 660
DB 601 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 660
QY 601 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 660
DB 601 CCGGACACAGGCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 660
QY 661 TCGGCTCTGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 720
DB 661 TCGGCTCTGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 720
QY 721 TCGGCTCTGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 780
DB 721 TCGGCTCTGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 780
QY 761 ACCGACACCGGCGGACGAGATCTCAACGAGATTCGAGGACACACTGATATCT 840
DB 761 ACCGACACCGGCGGACGAGATCTCAACGAGATTCGAGGACACACTGATATCT 840

```

```

QY 781 CCTACCACTCTTGGCAGGCAATAGACCAAGACTGATGCAATGACCTACCATAGCAG 840
DB 841 CCTACCACTCTTGGCAGGCAATAGACCAAGACTGATGCAATGACCTACCATAGCAG 900
QY 841 GTACTCTCTCACTCTCTTCAATCAAGCACTTCTCCCAAGTTGCTCACTGGGCTCT 900
DB 901 GTACTCTCTCACTCTCTTCAATCAAGCACTTCTCCCAAGTTGCTCACTGGGCTCT 960
QY 901 TTTCTTTCTGCTCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 960
DB 961 TTTCTTTCTGCTCTTTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 1020
QY 961 AGACCCGACTACTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTTTCAGATTTAT 1020
DB 1021 AGACCCGACTACTCAAGAGCTGCAAGAGACATTTCTGAATGTTTTTTCAGATTTAT 1080
QY 1021 AAAAAGAGGGGGTTTTTCTGGGCTCTCAATATTTAAGTTCAAGCCAGATCTGTGTGTA 1080
DB 1081 AAAAAGAGGGGGTTTTTCTGGGCTCTCAATATTTAAGTTCAAGCCAGATCTGTGTGTA 1140
QY 1081 CAATGACTCTGGGCTTCCGAAAGGATCAATCAATGTCACAGCTGAGACAGCTTC 1140
DB 1141 CAATGACTCTGGGCTTCCGAAAGGATCAATCAATGTCACAGCTGAGACAGCTTC 1200
QY 1141 AATCAGTATTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1201 AATCAGTATTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1201 AGTCATGTCATTTCTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 1261 AGTCATGTCATTTCTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1261 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1321 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 1381 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
QY 1381 ACCTACATCTATAGAGAGAGATCCCACTTACACACACACACACACACACACAC 1440
DB 1441 ACCTACATCTATAGAGAGATCCCACTTACACACACACACACACACACACAC 1500
QY 1441 AGCAGTACCGATCTGATGCTTCTGAGAGAGATTTCTGAGAGATTTGAGAGAG 1500
DB 1501 AGCAGTACCGATCTGATGCTTCTGAGAGAGATTTCTGAGAGATTTGAGAGAG 1560
QY 1501 TCTTACCAAAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1561 TCTTACCAAAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1561 TGAAGTGAATGG 1572
DB 1621 TGAAGTGAATGG 1632

```

RESULT 4

HSTEXMA

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLE

JOURNAL

HSTEXMA 1721 bp mRNA linear PRI 14-AUG-1995
Human mRNA for transmembrane epithelial tumor mucin antigen.
X52229.1 GI:37053
transmembrane protein; tumor antigen.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1721)
Wreschner D.H.
Direct Submission
Submitted (16-MAR-1990) Wreschner D.H., Tel Aviv University, Dept
of Microbiology, Faculty of Life Sciences, Tel Aviv 69978, Israel

REFERENCE 2 (bases 1 to 1721)
Wreschener, D.H., Hareuveni, M., Tsarfay, I., Smorodinsky, N.,
Horov, J., Zaretsky, J., Kokes, P., Weiss, M., Lache, R., Dion, A.S. and
Keydar, I.
TITLE Human epithelial tumor antigen cDNA sequences. Differential
splicing may generate multiple protein forms.
JOURNAL Eur. J. Biochem. 189 (3), 463-473 (1990).
MEDLINE 90276413
PUBMED 2351132
REFERENCE 3 (bases 1 to 1721)
Hareuveni, M., Tsarfay, I., Zaretsky, J., Kokes, P., Horov, J.,
Zrihan, S., Weiss, M., Green, S., Lache, R., Keydar, I. and
Wreschener, D.H.
TITLE A transcribed gene, containing a variable number of tandem repeats,
codes for a human epithelial tumor antigen. cDNA Cloning,
expression of the transcribed gene and over-expression in breast
cancer tissue
JOURNAL Eur. J. Biochem. 189 (3), 475-486 (1990)
MEDLINE 90276414
PUBMED 2112460
COMMENT See <X52228> for secreted tumour antigen.
FEATURES
Source Location/Qualifiers
1..1721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1q21-24"
/cell_line="T47D breast tumour"
/clone_lib="lambda gtl1"
58..1605
/codon_start=1
/product="precursor polypeptide (AA -21 to 494)"
/protein_id="CAA36478.1"
/db_xref="GI:37054"
/db_xref="SWISS-PROT:P15941"
/analysis="MTPQSPPLLLTLVTVVYVGSAGHSSPTGGEKETSATORSS
VSSSTEKNAVSMSSVLSHSPGSSSTGGDVTILAPTPPAGSAATMGDVTSP
VTRPALGSTPPADVTASAPDNKAPAGSTAPPAHVTSAPDTPRSTAPPAHVTS
ADPTPPGSTAPPAHVTSAPDTPRSTAPPAHVTSAPDTPRSTAPPAHVTS
TSAGSAGSAGSSTLVHNGTSARATTPPASKTSPFSPHSHSDPTTLASHSTKTPASS
THSTVPPPLTSSNHSSTPOLSTSVSPFSLHSLQNSLSDPSTDYVDELORDIS
EMFLDIYAGGFLGLSNIKRPGSVVQTLTLFREGTINHDVETQFNQYKELASRY
NLTJSDVSDVPPFPAQAGVPMGJALIVCLVALAIYLIALLAVCCRRKN
YQGLDIPPARDTYHPMSEYPTHTHGRVYPPSTDRSPYKVSAGNGSSLSYTPAV
AATSNL."
sig_peptide 58..120
mat_peptide 121..1602
BASE COUNT 339 a 634 c 395 g 353 t
ORIGIN
Query Match 94.2%; Score 1481.2; DB 9; Length 1721;
Best Local Similarity 95.5%; Pred. No. 1.8e-260;
Matches 1559; Conservative 0; Mismatches 13; Indels 60; Gaps 1;

QY 301 CCGGACACGGAAACAGCTTCAAGTTCACTGCCACTGGGACACGATGTCACCTGGCTC 360
DB 301 CCGGACACGGAAACAGCTTCAAGTTCACTGGCTGCCACTGGGACACGATGTCACCTGGCTC 360
QY 361 CCACTCAGGAGCCAGCCCTGGGCTCCACACCCGACGACCCAGTGTCACTCAGGCC 420
DB 361 CCACTCAGGAGCCAGCCCTGGGCTCCACACCCGACGACCCAGTGTCACTCAGGCC 420
QY 421 CCGGACAAAGAGCCAGCCCTGGGCTCCACACCCGACGACCCAGTGTCACTCAGGCC 480
DB 421 CCGGACAAAGAGCCAGCCCTGGGCTCCACACCCGACGACCCAGTGTCACTCAGGCC 480
QY 481 CCGGACACGAGCCGCCCCGAGCTCCACCGCC-----CCGCCCCCAGCCAGTGTCACTCAGGCC 513
DB 481 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 540
QY 514 -----CCGCCCCCAGCCAGTGTCACTCAGGCC 540
DB 541 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 600
QY 541 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 600
DB 601 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 660
QY 601 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 660
DB 661 CCGGACACGAGCCGCCCCGAGCTCCACCGCCCGCGACCGGTCACCTCAGGCC 720
QY 661 TCAGGCTGTGATCAGGCTCAGCTTCTACTGTGTGACAAAGGACCTTGCAGAGGCT 720
DB 721 TCAGGCTGTGATCAGGCTCAGCTTCTACTGTGTGACAAAGGACCTTGCAGAGGCT 780
QY 721 ACCACACCCAGCCAGCAGACACTCCACCCAGACATTCGACCCACCTCTGTACT 780
DB 781 ACCACACCCAGCCAGCAGACACTCCACCTCTGTACTCTGTACT 840
QY 781 CCTACACCTTGCACAGCATAGACCAAGACCTGATGACATGACATGACATGACAG 840
DB 841 CCTACACCTTGCACAGCATAGACCAAGACCTGATGACATGACATGACATGACAG 900
QY 841 GATACCTCTCAGCTCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 900
DB 901 GATACCTCTCAGCTCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 960
QY 901 TCTTTTCTCTGCTTTTCAATTTCAACTTCAGTTAATTCCTCTCTGGAAGATCCC 960
DB 961 TCTTTTCTCTGCTTTTCAATTTCAACTTCAGTTAATTCCTCTCTGGAAGATCCC 1020
QY 961 AGCAGCAGCTACTACAGAGGCTGAGAGAGCAATTCGAAATGTTTTTTCAGATTAT 1020
DB 1021 AGCAGCAGCTACTACAGAGGCTGAGAGAGCAATTCGAAATGTTTTTTCAGATTAT 1080
QY 1021 AAACAAAGGGGTTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGCCAGGATCTGTGTGTA 1080
DB 1081 AAACAAAGGGGTTTTTCTGGGCTCTCCAAATTAATTAAGTTCAGGCCAGGATCTGTGTGTA 1140
QY 1081 CAATTGACTCTGGGCTTCGAGAGAGGTACCAATGTCACAGACGTGAGACAGATTC 1140
DB 1141 CAATTGACTCTGGGCTTCGAGAGAGGTACCAATGTCACAGACGTGAGACAGATTC 1200
QY 1141 AATCAGTATTAAGAGAGAGAGCTCTCGATATTAACCTGACGATCTCAGACGTGAG 1200
DB 1201 AATCAGTATTAAGAGAGAGAGCTCTCGATATTAACCTGACGATCTCAGACGTGAG 1260
QY 1201 AGTCATGAGCAATTCCTTTCTCTGCTCCAGCTGAGGAGCTGAGGAGTCAAGCTGGGCAATC 1260
DB 1261 AGTCATGAGCAATTCCTTTCTCTGCTCCAGCTGAGGAGCTGAGGAGTCAAGCTGGGCAATC 1320
QY 1261 GGGCTGCTGGTCTGGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCT 1320
DB 1321 GGGCTGCTGGTCTGGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCTGTGTCT 1380

| | | | |
|----|------|--|------|
| Qy | 1321 | GCCTCTGTGAGTGCCTCCGCGAAAGAACTTACCGGGCAGCTGGACATCTTTTCAGGCCGGAGAT | 1380 |
| Dp | 1381 | GGGTCTGTGACGTGCTCCGCGAAAGAACTTACGGGGCAGCTGGACATCTTTTCAGGCCGGAGAT | 1440 |
| Qy | 1381 | ACCTACACACCTCTATGACGACGAGTACCCGACCTTACACACGCCATGGGGGGCTATGTGCCCCCT | 1440 |
| Dp | 1441 | ACCTACACCTCTATGACGACGAGTACCCGACCTTACACACGCCATGGGGGGCTATATGTGCCCCCT | 1500 |
| Qy | 1441 | AGCAGTACCGATTCGTACCCCTATGAGAAAGGTTCTGACAGTAAATGTGTGGAGAGAGCTTC | 1500 |
| Dp | 1501 | AGCAGTACCGATTCGTACCCCTATGAGAAAGGTTCTCAGAGTAAATGTGTGGAGAGAGCTTC | 1560 |
| Qy | 1501 | TCTTTACACAAACCCAGCAGATGGAGAGCCATTTCTGCAACTTTGTAAGGGGCAAGTGGCCCTC | 1560 |
| Dp | 1561 | TCTTTACACAAACCCAGCAGATGGAGAGCCATTTCTGCAACTTTGTAAGGGGCAAGTGGCCCTC | 1620 |
| Qy | 1561 | TGAGCTGAGTGG | 1572 |
| Dp | 1621 | TGAGCTGAGTGG | 1632 |

| | | | |
|------------|--------------------------------------|--------|-----------------|
| RESULT 5 | | | |
| AX35367 | | | |
| LOCUS | 1804 bp | DNA | |
| DEFINITION | Sequence 5876 from Patent WO0194629. | linear | PAT 09-JAN-2002 |

| | | |
|----------|--------------|-------------|
| VERSION | AX35367.1 | GI:18126086 |
| KEYWORDS | . | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |

REFERENCE
AUTHORS
1. Mammalia; Eutheria; Primates; Catarrhini, Hominoidea; Homo.
Young, P.E., Augustus M., Carter, K.C., Ebner, R., Endress, G.,
Hortigan, S., Soppet, D.R. and Weaver, Z.

gene sets
Journal Patent: WO 0194629-A 5876 13-DEC-2001;
Avalon Pharmaceuticals (US)
Features Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT      361 a      622 c      434 g      387 t
ORIGIN

```

Query Match 89.8%; Score 1412.2; DB 6; Length 1804;
Best Local Similarity 95.3%; Pred. NC. 7e-248;
Matches 1490; Conservative 0; Mismatches 13; Indels 60; Gaps 1.

[illegible]

Db 85 ACCAGCTCTCTTCTTCTGCTGCTCTCCACAGTGGTTACAGTTGACAGGTTCT 144

Qy 130 GGTTCATGCAAGCTTACCCCAAGGTTGAGAAAGAGACCTTGGGCTACCCAGAGAACTTCA 189

DB GGATGATGCAAGATCTTACATCCCAAGTGGAGAGAGAGAGATCTTGGGATTCACCAAGAGAGATCTTC 249

DY 190 GTGGCCCAAGCTTACTGAGAGAGAGATGCTGTGAGATATGGACCAAGCAAGCTTCTTCCAGACAC 249

DH 205 GTGCGCCAGGCTTCTTATGACAGATATGCTGTGAGATATGGACCAAGCAAGCTTCTTCCAGACAC 264

QY 250 AGCCCCGTTCAAGGCTCCTCCACCATCGAGGACAGATGTCACTTGGCCCCGGCACAG 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 265 AGCCCCGTTCAAGGCTCCTCCACCATCGAGGACAGATGTCACTTGGCCCCGGCACAG 324

310 GAACACACCTTCAGGTTACGTCGACACTGGGGACAGATGTCACCTTGGTCCACATCC 365
325 GAACCACTTCAGGTTACGTCGACACTGGGGACAGATGTCACCTTGGTCCACATCC 384

| | | | | | |
|----|-----|--------------------------|-----------------|-------------------------|-----|
| QY | 370 | AAGCCAGAGCCTTG6GCTCCACCA | CCCCCGCCAGGCCCA | CGATGTCACTTGAGCCCGGACAC | 429 |
| D6 | 385 | AGGCCAGAGCCTTG6GCTCCACCA | CCCCCGCCAGGCCCA | CGATGTCACTTGAGCCCGGACAC | 444 |
| QY | 430 | AAGCCAGAGCCTTG6GCTCCAC | CCCCCGCCCGCCCA | CGGTGTCACTTG6GCGCGGACAC | 489 |
| D6 | 445 | AAGCCAGAGCCTTG6GCTCCAC | CCCCCGCCCGCCCA | CGGTGTCACTTG6GCGCGGACAC | 474 |

| | | | |
|----|-----|--|-----|
| QY | 490 | AAGCCGCCCCCGGAGCTCAACGCCCCCGCCGCGCCGCGAGTGCATCTGCGGCCGACAC | 549 |
| | | | |
| | 475 | -----GCCACGGTGTCACTCTGCGCCCGGACACC | 504 |
| Db | | | |

D6
D7

505 AGGCGGGCCCGGAGCTTCACCGCCCCCCCAGCCCTGGTGTCACTTGAGGCCCGGACAAC 564

610 AAGCGGAGCTTTGGGCTCACCGGCCCCCTCCAGTGCACAAATGTCACTCGAGCTCAGGCTCT 668

D6 565 AAGCCCGCGCTTTGGGCTCCACCGCCCCCTCCAGTCAAAATGTACATCGAGGCTCAGGCTCT 62

D7 670 GCATCAGGCTTCAGCTTCTATCTGTGGTGCACACGCACTTCGCAAGGCTACCAAC 72

QY 730 CCAGCAGCAGAGCACTCCACCAGCATTTCCAGGCACACTCTGATCTCTACACC 7898
Db 685 CCAGCAGCAGAGCACTCTCATTTCCATTTCCAGGCACACTCTGATCTCTACACC 744

QY 790 CTTGGCAACCATAGCAACAAAGCTGATGCGACAGTAGACCTGCACCATATACACAGGTAACCTCT 845
 |||||
 DB 745 CTTGGCAACCATAGCAACAAAGCTGATGCGACAGTAGACCTGCACCATATACACAGGTAACCTCT 804

| | | |
|-----|---|----|
| 87 | CTGCTCTTTCACATTTAAACCTCCAGATTAACTCTCTGGAGAGATCCAGACCGAC | 96 |
| D6 | CTGACCTCTCAATCAGAGCACTCTCCCAAGTGTCTACTGGGGTCTCTTCTTTTC | 86 |
| 910 | CTGCTCTTTCACATTTAAACCTCCAGATTAACTCTCTGGAGAGATCCAGACCGAC | 96 |

Db 865 CTGTCTTTTCACATTTTCAAACTCCAGTTAATCTCTCTGGAAGATCCCGACACCGAC 92

Oy 970 TACTACCAAGAGCTGCAGAGAGACATTTTGGAAATGTTTTGCGATTTTAAACAAGG 102

[illegible]

| Qy | 1090 | 1045 | 1105 |
|--|--|--|------|
| CTGGACCTTCGAGAGGTACATCAATGCCAGTCGTGGACACAGTTCAATCAGTAT | CTGGACCTTCGAGAGGTACATCAATGCCAGTCGTGGACACAGTTCAATCAGTAT | CTGGACCTTCGAGAGGTACATCAATGCCAGTCGTGGACACAGTTCAATCAGTAT | |

Dd 1105 AAAAGCGAAGACGCTTCGATATTAAGTCAATCTCAGAAGTCAGCGATGATATGG 116

Db 1165 CATTTCCTCTGCGCCAGCTGAGGGCTGAGGATCCAGGCTGGGGCATCGAGCTG 122

Gy 1270 GTCGTGCTGTGATTCGAGTGCAGCTGGCAATGTCATATCAATTCGCTTGGCTGTCTGT 133

| | | | |
|----|------|--|------|
| Db | 1225 | GTGCTGGTCTGTGTTCTGGTTGGGCGTGCACATGTCATCTCAATGACCTTGGGCTGTCTGT | 1226 |
| Qy | 1330 | CAGTCCGCGCGAAGAATCAACGGGACGCTGGACATCTTTTCCAGCCCGGGATATCAATCAAT | 1331 |

Qy 1390 CCTATAGCGAGTACCCCACTTACGACCCATGGGCGTATGTGCCCCCTACAGTACC 14
Db 1345 CCTATAGCGAGTACCCCACTTACGACCCATGGGCGTATGTGCCCCCTACAGTACC 14

| Qy | 1450 | GATCGTAGCCCTTATGAAGTTTCTGAGGTAATGGTGAGAGAGCTCTTTACACA | 1509 |
|----|------|--|------|
| Db | 1405 | GATCGTAGCCCTTATGAAGTTTCTGAGGTAATGGTGAGAGAGCTCTTTACACA | 1464 |
| Qy | 1510 | AATCCAGAGATGGAGGCACTTCTGCCACTTTGATGGGACAGCTGCCCTCTGAAGCTAG | 1569 |
| Db | 1465 | AATCCAGAGATGGAGGCACTTCTGCCACTTTGATGGGACAGCTGCCCTCTGAAGCTAG | 1524 |
| Qy | 1570 | TGG | 1572 |
| Db | 1525 | TGG | 1527 |

| | |
|------------|--|
| RESULT 6 | |
| HUMMUCAB | |
| LOCUS | HUMMUCAB |
| DEFINITION | 1804 bp |
| ACCESSION | human polymorphic epithelial mucin (PEM) |
| J05581 | mRNA, complete cds. |
| | PRI 07-JAN-1995 |

| | | |
|----------|---|-----------|
| VERSION | J05581.1 | GI:188869 |
| KEYWORDS | polymorphic epithelial mucin. | |
| SOURCE | Homo sapiens adult adenocarcinoma CDNA to mRNA. | |
| ORGANISM | Homo sapiens | |

REFERENCE
AUTHORS

1 (bases 1 to 1804)
Gendler, S.U., Lancaster, C.A., Taylor-Padgett, J., Dubig, T., Peat, N., Burchell, J., Pemberton, J., Talani, R. and Wilson, B.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

| TITLE | Author(s) | JOURNAL | Volume | Page(s) | Year |
|---|----------------------|----------------|----------|-------------|--------|
| Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin | Watanabe, T., et al. | J. Biol. Chem. | 265 (25) | 15286-15293 | (1990) |
| Journal of Molecular Medicine | | | 90368715 | | |
| DRUMPD | | | | | |
| 10000000 | | | | | |

| | |
|----------|---|
| FORMALIN | 167/389 |
| COMMENT | Draft entry and computer-readable sequence for [J. Biol. Chem. (1990) In press] kindly submitted by S.J.Gendler, 26-JUN-1990. |
| FEATURES | Location/Qualifiers |

```
source
1. 1804
/organism="Homo sapiens"
/db xref="taxon:9606"
/map="Xq26.3-q27.1"
/contig="1804-1805-1806-1807-1808-1809-1810-1811-1812-1813-1814-1815-1816-1817-1818-1819-1820-1821-1822-1823-1824-1825-1826-1827-1828-1829-1830-1831-1832-1833-1834-1835-1836-1837-1838-1839-1840-1841-1842-1843-1844-1845-1846-1847-1848-1849-1850-1851-1852-1853-1854-1855-1856-1857-1858-1859-1860-1861-1862-1863-1864-1865-1866-1867-1868-1869-1870-1871-1872-1873-1874-1875-1876-1877-1878-1879-1880-1881-1882-1883-1884-1885-1886-1887-1888-1889-1890-1891-1892-1893-1894-1895-1896-1897-1898-1899-1900-1901-1902-1903-1904-1905-1906-1907-1908-1909-1910-1911-1912-1913-1914-1915-1916-1917-1918-1919-1920-1921-1922-1923-1924-1925-1926-1927-1928-1929-1930-1931-1932-1933-1934-1935-1936-1937-1938-1939-1940-1941-1942-1943-1944-1945-1946-1947-1948-1949-1950-1951-1952-1953-1954-1955-1956-1957-1958-1959-1960-1961-1962-1963-1964-1965-1966-1967-1968-1969-1970-1971-1972-1973-1974-1975-1976-1977-1978-1979-1980-1981-1982-1983-1984-1985-1986-1987-1988-1989-1990-1991-1992-1993-1994-1995-1996-1997-1998-1999-2000-2001-2002-2003-2004-2005-2006-2007-2008-2009-2010-2011-2012-2013-2014-2015-2016-2017-2018-2019-2020-2021-2022-2023-2024-2025-2026-2027-2028-2029-2030-2031-2032-2033-2034-2035-2036-2037-2038-2039-2040-2041-2042-2043-2044-2045-2046-2047-2048-2049-2050-2051-2052-2053-2054-2055-2056-2057-2058-2059-2060-2061-2062-2063-2064-2065-2066-2067-2068-2069-2070-2071-2072-2073-2074-2075-2076-2077-2078-2079-2080-2081-2082-2083-2084-2085-2086-2087-2088-2089-2090-2091-2092-2093-2094-2095-2096-2097-2098-2099-2100-2101-2102-2103-2104-2105-2106-2107-2108-2109-2110-2111-2112-2113-2114-2115-2116-2117-2118-2119-2120-2121-2122-2123-2124-2125-2126-2127-2128-2129-2130-2131-2132-2133-2134-2135-2136-2137-2138-2139-2140-2141-2142-2143-2144-2145-2146-2147-2148-2149-2150-2151-2152-2153-2154-2155-2156-2157-2158-2159-2160-2161-2162-2163-2164-2165-2166-2167-2168-2169-2170-2171-2172-2173-2174-2175-2176-2177-2178-2179-2180-2181-2182-2183-2184-2185-2186-2187-2188-2189-2190-2191-2192-2193-2194-2195-2196-2197-2198-2199-2200-2201-2202-2203-2204-2205-2206-2207-2208-2209-2210-2211-2212-2213-2214-2215-2216-2217-2218-2219-2220-2221-2222-2223-2224-2225-2226-2227-2228-2229-2230-2231-2232-2233-2234-2235-2236-2237-2238-2239-2240-2241-2242-2243-2244-2245-2246-2247-2248-2249-2250-2251-2252-2253-2254-2255-2256-2257-2258-2259-2260-2261-2262-2263-2264-2265-2266-2267-2268-2269-2270-2271-2272-2273-2274-2275-2276-2277-2278-2279-2280-2281-2282-2283-2284-2285-2286-2287-2288-2289-2290-2291-2292-2293-2294-2295-2296-2297-2298-2299-2300-2301-2302-2303-2304-2305-2306-2307-2308-2309-2310-2311-2312-2313-2314-2315-2316-2317-2318-2319-2320-2321-2322-2323-2324-2325-2326-2327-2328-2329-2330-2331-2332-2333-2334-2335-2336-2337-2338-2339-2340-2341-2342-2343-2344-2345-2346-2347-2348-2349-2350-2351-2352-2353-2354-2355-2356-2357-2358-2359-2360-2361-2362-2363-2364-2365-2366-2367-2368-2369-2370-2371-2372-2373-2374-2375-2376-2377-2378-2379-2380-2381-2382-2383-2384-2385-2386-2387-2388-2389-2390-2391-2392-2393-2394-2395-2396-2397-2398-2399-2400-2401-2402-2403-2404-2405-2406-2407-2408-2409-2410-2411-2412-2413-2414-2415-2416-2417-2418-2419-2420-2421-2422-2423-2424-2425-2426-2427-2428-2429-2430-2431-2432-2433-2434-2435-2436-2437-2438-2439-2440-2441-2442-2443-2444-2445-2446-2447-2448-2449-2450-2451-2452-2453-2454-2455-2456-2457-2458-2459-2460-2461-2462-2463-2464-2465-2466-2467-2468-2469-2470-2471-2472-2473-2474-2475-2476-2477-2478-2479-2480-2481-2482-2483-2484-2485-2486-2487-2488-2489-2490-2491-2492-2493-2494-2495-2496-2497-2498-2499-2500-2501-2502-2503-2504-2505-2506-2507-2508-2509-2510-2511-2512-2513-2514-2515-2516-2517-2518-2519-2520-2521-2522-2523-2524-2525-2526-2527-2528-2529-2530-2531-2532-2533-2534-2535-2536-2537-2538-2539-2540-2541-2542-2543-2544-2545-2546-2547-2548-2549-2550-2551-2552-2553-2554-2555-2556-2557-2558-2559-2560-2561-2562-2563-2564-2565-2566-2567-2568-2569-2570-2571-2572-2573-2574-2575-2576-2577-2578-2579-2580-2581-2582-2583-2584-2585-2586-2587-2588-2589-2590-2591-2592-2593-2594-2595-2596-2597-2598-2599-2600-2601-2602-2603-2604-2605-2606-2607-2608-2609-2610-2611-26
```

```

gene
      /cell_line="D120"
      /tissue_type="adenocarcinoma"
      /dev_stage="adult"
      1..1804
      /gene="MCNA"

```

```

mRNA      <1..1804
           /gene="MCNA"
           /product="PEM mRNA"
CDS        73..1500
           /gene="MCNA"

```

```

/ gene = "MUCNA"
/ codon_start = 1
/ product = "polymorphic epithelial mucin"
/ protein_id = "AAAS9876.1"
/ db_xref = "GI:188870"

```

[illegible]

SLEPESDIPYQLOQRDI SEMFQIYKQGGJAGSINTKFRPSSVQVTLFARSTENQ
 HVEFONROYKTEAASRNVLTISDVSVDVFPFSSQSGAGVPGMGIALMLVQVIVA
 LAIVLIVLAVOCQRKRYGQLDIPFADYTHPMSEIPIYTHHGRVYPSSTDRSPE
 KYKSGAGGSSLSITNPAVATSAANI."

| | | |
|-------------|-----------|---|
| sig_peptide | 73..135 | /gene="MCNA" |
| mat_peptide | 136..1497 | /gene="MCNA" |
| | | /product="holymorphic antih[?]a] mucin" |

| polya_signal | 1783 . 1788 | /gene="MCNAA" |
|--------------|-------------|-------------------|
| BASE COUNT | 361 a | 622 c 434 g 387 t |
| ORIGIN | | |

| | | | | |
|-----------------------|-----------------|------------------|------------|--------------|
| Query Match | 89.8% | Score 1412.2; | DB 9; | Length 1804; |
| Best Local Similarity | 95.3% | Pred. No. 7e248; | | |
| Matches 1490; | Conservative 0; | Mismatches 13; | Indels 60; | Gaps 1; |

OY 10 GGGCGTGAATCTGTTTCCGCCCTCCCAACCAATTTCACCAACCACTGACACACGGG 69
 Db 25 GCTGCGTAATCTGTTCTGCCCTCTCCCAACCAATTTCACCAACCACTGACACACGGG 84
 OY 70 ACCGAGTCTCTTTCTTCTGCTGCTGCTCTCAAGTGTTCACAGTTTACAGTTCT 129
 Db 85 ACCAGTCTCTTCTTCTGCTGCTGCTCTCAAGTGTTCACAGTTTACAGTTCT 144
 OY 130 GGTGATCAAGCTCTACCCAGGTGAGAAAGAGACTTCGGCTACCCAGAGATTCA 189
 Db 145 GGTGATCAAGCTCTACCCAGGTGAGAAAGAGACTTCGGCTACCCAGAGATTCA 204
 OY 190 GTGCGCAGCTCTACGTGAGAAAGATGCTGTGATGACAGAGGTACTCTCCAGCAC 249
 Db 205 GTGCGCAGCTCTACGTGAGAAAGATGCTGTGATGACAGAGGTACTCTCCAGCAC 264
 OY 250 AGCCCGGTTCAGGCTCTCTCCACACTTCAGGGAAGAGATGTCACTCTGCGCCCGCACG 309
 Db 265 AGCCCGGTTCAGGCTCTCTCTCCACACTTCAGGGAAGAGATGTCACTCTGCGCCCGCACG 324
 OY 310 GAACAGACTTCAGGTTCAAGCTGACCTGAGCCACCGGGGACAGGATGTCACTCGGTCCAGTACC 365
 Db 325 GAACAGACTTCAGGTTCAAGCTGACCTGAGCCACCGGGGACAGGATGTCACTCGGTCCAGTACC 384
 OY 370 AGGCAAGCTTCAGGCTCTCCACCAACCCGCGACGCTCAAGATGTCACTCAAGCCCGACAAC 429
 Db 385 AGGCAAGCTTCAGGCTCTCCACCAACCCGCGACGCTCAAGATGTCACTCAAGCCCGACAAC 444
 OY 430 AAGCAAGCTTCAGGCTCTCCACCGCCCCCGACGAGTGTCACTCGGCCCCCGACAAC 489
 Db 445 AAGCAAGCTTCAGGCTCTCCACCGCCCCCGACGAGTGTCACTCGGCCCCCGACAAC 474
 OY 490 AGGCGCGCCCGGGCTCCACCGCCCCCGGCGCCAGAGTGTCACTCGGCCCCCGACAAC 549
 Db 475 -----GCCACGAGTGTCACTCGGCCCCCGACAAC 504
 OY 550 AGGCGCGCCCGGGCTCCACCGCCCCCGGCGCCAGAGTGTCACTCGGCCCCCGACAAC 609
 Db 505 AGGCGCGCCCGGGCTCCACCGCCCCCGGCGCCAGAGTGTCACTCGGCCCCCGACAAC 564
 OY 610 AGGCGCGCTTGGGCTCCACCGCCCCCGCTCAAGTGTCACTCGGCCCCCGACAAC 669
 Db 565 AGGCGCGCTTGGGCTCCACCGCCCCCGCTCAAGTGTCACTCGGCCCCCGACAAC 624
 OY 670 GCATCAAGGCTGACTTCTACTGTGTGACAAAGGCACTCTGCGAGGCTTACCAACAC 729
 Db 625 GCATCAAGGCTGACTTCTACTGTGTGACAAAGGCACTCTGCGAGGCTTACCAACAC 684
 OY 730 CGAGCCAGCAAGAGCACTCCACCAGATTCGCGACCACTGTGATCTCTCAACAC 789
 Db 685 CGAGCCAGCAAGAGCACTCCACTTCTCAATTCCAGCACTGTGATCTCTCAACAC 744
 OY 790 CTTCGACGCAATGACCAAGACGATGACGATGACCACTCAATGACGATGACCTCTCT 849
 Db 745 CTTCGACGCAATGACCAAGACGATGACGATGACCACTCAATGACGATGACCTCTCT 804
 OY 850 CTCACTCTCTCAATCAAGCACTTCTCCCAAGTGTCTACTGGGCTCTTCTTCTTTTC 909
 Db 805 CTCACTCTCTCAATCAAGCACTTCTCCCAAGTGTCTACTGGGCTCTTCTTCTTTTC 864
 OY 910 CTGTCTTTTCAATTTCAAAACCTCAAGTTAATTCCTCTGGAAGATCCCAAGACCGAC 969
 Db 865 CTGTCTTTTCAATTTCAAAACCTCAAGTTAATTCCTCTGGAAGATCCCAAGACCGAC 924
 OY 970 TACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTGACAGATTTTAAACAAGG 1029
 Db 925 TACTACCAAGAGCTGACAGAGACATTTCTGAAATGTTTTTGACAGATTTTAAACAAGG 984
 OY 1030 GGTGTTCTGGGCTCTCCAAATATTAAGTTCAAGGCCAGATCTGTGATGTACAAATTGACT 1089

|||||
Db 985 GGTTCCTGGGCTCTCCAAATATTAAGTTGAGGCGAGATCTGTGGTGAATAATGACT 1044
|||
Qy 1090 CTGGCCCTTCGAGAAAGAGTACATTCATAGTCCAGACAGTGAAGACACAGTTCAATCAGAT 1149
|||
Db 1045 CTGGCTCTCCGAAAGAGTACATTCATAGTCCAGACAGTGAAGACACAGTTCAATCAGAT 1104
|||
Qy 1150 AAAACGAGAACAGCTCTGATATTAACCTGACGATCTCAGACGCTGAGGTGATGATG 1209
|||
Db 1105 AAAACGAGAACAGCTCTGATATTAACCTGACGATCTCAGACGCTGAGGTGATGATG 1164
|||
Qy 1210 CCATTTCCTTCCTGCGGAGTCTGAGGCTGAGGATGCGAGCTGAGGAGCATGCGCTGCTG 1269
|||
Db 1165 CCATTTCCTTCCTGCGGAGTCTGAGGCTGAGGATGCGAGCTGAGGAGCATGCGCTGCTG 1224
|||
Qy 1270 GTGCTGTCTGTGTCTGTGTGTGCGCTGAGCATGTCTATCTCATTCGCTTGGCTGTCTGT 1329
|||
Db 1225 GTGCTGTCTGTGTCTGTGTGTGCGCTGAGCATGTCTATCTCATTCGCTTGGCTGTCTGT 1284
|||
Qy 1330 CAGTGCCTGCGGAAAGAACTACGAGGCGAGCTGAGCATCTTCACGCGGAGTACCTACCAT 1389
|||
Db 1285 CAGTGCCTGCGGAAAGAACTACGAGGCGAGCTGAGCATCTTCACGCGGAGTACCTACCAT 1344
|||
Qy 1390 CCTATGAGCGAGTACCCCACTACCAACCCAGGCGCTATGTGCCCTTACGAGTACC 1449
|||
Db 1345 CCTATGAGCGAGTACCCCACTACCAACCCAGGCGCTATGTGCCCTTACGAGTACC 1404
|||
Qy 1450 GATGCTAGCCCTTATGAGAGGTTTCTGAGGATATGATGAGGAGCAGACCTCTTACACA 1509
|||
Db 1405 GATGCTAGCCCTTATGAGAGGTTTCTGAGGATATGATGAGGAGCAGACCTCTTACACA 1464
|||
Qy 1510 AAACCCAGAGTGGAGGACCACTTCTGCAACTTTGTAAGGGGACGTCGCTCTGAGCTGAG 1569
|||
Db 1465 AAACCCAGAGTGGAGGACCACTTCTGCAACTTTGTAAGGGGACGTCGCTCTGAGCTGAG 1524
|||
Qy 1570 TGG 1572
|||
Db 1525 TGG 1527

RESULT 7
GIBMUC1A 1795 bp DNA linear PRI 17-JUL-1995
LOCUS Hylobates lar epithelial mucin (MUC1) gene, complete cds.
DEFINITION
ACCESSION L41589
VERSION L41589.1 GI:897611
KEYWORDS MUC1 gene; epithelial mucin; membrane protein; mucin; variable number tandem repeat.
SOURCE Hylobates lar (clone library: cos203 cosmid library) adult DNA.
ORGANISM Hylobates lar
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
REFERENCE 1 (bases 1 to 1795)
AUTHORS Spicer, A.P., Dubig, T., Chilton, B.S. and Gendler, S.J.
TITLES Analysis of mammalian MUC1 genes reveals potential functionally important domains
JOURNAL Genomics (1995) In press
COMMENT On Jul 14, 1995 this sequence version replaced gi:886240.
The gibbon sequence presented here contains two consensus 60 bp tandem repeat units. From studies of the human MUC1 gene, a single allele can contain from 20 up to 125 repeat units. The gibbon MUC1 clones we describe here contained approximately 35 tandem repeats.
FEATURES
source
1..1795
/organism="Hylobates lar"
/db_xref="taxon:9580"
/clone_lib="cos203 cosmid library"
/dev_stage="adult"
1..1795
/gene="MUC1"
1..1795
/gene="MUC1"
/product="epithelial mucin"

CDS
69..1496
/gene="MUC1"
/note="1287..1493 cytoplasmic domain, 1194..1286 transmembrane domain"
/codon_start=1
/product="epithelial mucin"
/protein_id="AA66965.1"
/db_xref="GI:897612"
/translation="MTPTQSLFLLILLTVLVTVVSGHASTPGEKETSATRRSS
MESETRKYVNTSSVLSHSRSGSSSTTGODVSLAPETPAGSAATWGDVTP
VIRPAPGSTSPADVTSAPTRPALGSTAPYKHVTSAPDTRPLGSTAPYHGVTS
ADPTRTPLGSTAPYHNVTSASGSASTLVHNGTSARATTPAKSTPESIPSHH
SDPTPLTSHSTKTDASTHSTVSPLTSHNSTPQLSTVFSFPLSHLNRS
SLDPTNYOELORDISELLIQYKQGBFLSNIKRPSVVQSTLAFRGTTNV
HDVEAFNQHKEAASRYNLISDVSVDVPPFPFSAQSGADVPMGIALVLVQCVVA
LATVLAALAVCCRRKRYGQLDIFPARDAVHPMSEYPTVTHGRVYVPSPTWRSPE
KYSBNGSGSLSYINPAVATSNLV"
69..131
/gene="MUC1"
sig_peptide
453..572
/repeat_region
/rpt_type=tandem
/rpt_unit=453..512
repeat_unit
453..512
/gene="MUC1"
repeat_unit
513..572
/gene="MUC1"
polya_signal
1774..1779
/gene="MUC1"
BASE COUNT 365 a 429 g 390 t
ORIGIN
Query Match 84.5%; Score 1328.4; DB 9; Length 1795;
Best Local Similarity 91.9%; Pred. No. 1.4e-232;
Matches 1438; Conservative 0; Mismatches 66; Indels 60; Gaps 1;
9 TGAGCTGCTTAATCTGTTCTGCCCCCTCCCAACCATTTTACACACACACACACCGG 68
Db 20 TGCTTCCTGAGTCTGTTCTGCCCCCTCCCAACCATTTTACACACACACACACCGG 79
|||
Qy 69 CACCCAGTCTCTCTTCCTCCGAGGCTGCTCTCTCAAGTCTCAAGTTTCAAGTTTC 128
|||
Db 80 CACCCAGTCTCTCTTCCTCCGAGGCTGCTCTCTCAAGTCTCAAGTTTCAAGTTTC 139
|||
Qy 129 TGCTCATGCAAGCTTACCCCAAGTGAAGAAAGAAAGAAAGAAAGAAAGTTTC 188
|||
Db 140 CGGTCAATGCAAGCTTACCCCAAGTGAAGAAAGAAAGAAAGAAAGAAAGTTTC 199
|||
Qy 189' AGTGCACAGCTCTACTGAGAAAGAAAGTGTGAGTATGACACAGAGCTACTCTCCAGCA 248
|||
Db 200 AATGCCAAGCTTACTGAGAAAGAAAGTGTGAGTATGACACAGAGTATCTCCAGCA 259
|||
Qy 249 CAGCCCCGGTTAGGCTCTCTCCACACTCAGGACAGATGTCACTTGGCCCCCGGAC 308
|||
Db 260 CAGCCCCGGTTAGGCTCTCTCCACACTCAGGACAGATGTCACTTGGCCCCCGGAC 319
|||
Qy 309 GGAACAGCTTCAAGTGTCAAGTGTCACTGAGGACAGATGTCACTTGGCCCCCGGAC 368
|||
Db 320 GGAACAGCTTCAAGTGTCAAGTGTCACTGAGGACAGATGTCACTTGGCCCCCGGAC 379
|||
Qy 369 CAGGCCAGCTTGGGCTCTCCACACCCGCGACAGATGTCACTTGGCCCCCGGAC 428
|||
Db 380 CAGGCCAGCTTGGGCTCTCCACACCCGCGACAGATGTCACTTGGCCCCCGGAC 428
|||
Qy 429 CAGGCCAGCTTGGGCTCTCCACACCCGCGCGCCACGAGTGTCACTTGGCCCCCGGAC 488
|||
Db 429 -----GCCCGGACAC 439
|||
Qy 489 CAGGCCAGCTTGGGCTCTCCACACCCGCGCGCCACGAGTGTCACTTGGCCCCCGGAC 548
|||
Db 440 CAGGCCAGCTTGGGCTCTCCACACCCGCGCGCCACGAGTGTCACTTGGCCCCCGGAC 499
|||
Qy 549 CAGGCCAGCTTGGGCTCTCCACACCCGCGCGCCACGAGTGTCACTTGGCCCCCGGAC 608

```
|||||
Db 500 CAGGCCACCTTGGGCTCCACCGCCCCCAGTCCAGGATGCTACCTCGGCCCCGAGCAC 559
Qy 609 CAGGCCGCGCTTGGGCTCCACCGCCCCCAGTCCAGGATGCTACCTCGGCCCCGAGCTC 668
Db 560 CAGGCCACCTTGGGCTCCACCGCCCCCAGTCCAGGATGCTACCTCGGCCCCGAGCTC 619
Qy 669 TGCATCAGGCTCAGCTTCTACTGTGTGCAACCGGCACTCTGCGAGGGCTACCAAC 728
Db 620 TGCATCAGGCTCAGCTTCTACTGTGTGCAACCGGCACTCTGCGAGGGCTACCAAC 679
Qy 729 CCGACGCAAGAGAGAGCTCCACCGAGATTCGACGCGACCACTGTGATGCTCCACAC 788
Db 680 CCGACGCAAGAGAGAGCTCCACCGAGATTCGATTCGACGCGACCACTGTGATGCTCCAC 739
Qy 789 CCGTTCAGGCTTACGACCAAGATGATGCTGAGTACCTCACTCACTGATGCTCC 848
Db 740 CCGTTCAGGCTTACGACCAAGATGATGCTGAGTACCTCACTCACTGATGCTCC 799
Qy 849 TCTCACCTCCCTCACTCAAGAGCTTCTCCAGTGTCTAAGGAGTCTCTTCTTTT 908
Db 800 TCTCACCTCCCTCACTCAAGAGCTTCTCCAGTGTCTAAGGAGTCTCTTCTTTT 859
Qy 909 CCGTCTCTTTCATTCATTCAACTCCAGTTCATTCCTCTCTGGAAGATCCAGACCGA 968
Db 860 CCGTCTCTTTCATTCATTCAACTCCAGTTCATTCCTCTCTGGAAGATCCAGACCGA 919
Qy 969 CTATCACCAAGAGCTGACAGAGACATTTCTGGAAGTTCCTGACATTTATTAACAAG 1028
Db 920 CTATCACCAAGAGCTGACAGAGACATTTCTGGAAGTTCCTGACATTTATTAACAAG 979
Qy 1029 GAGTTTCTGGGCTCTCCAAATATTAAGTTCAGGCGAGATCTGTGGTGGTCAATGAC 1088
Db 980 GAGTTTCTGGGCTCTCCAAATATTAAGTTCAGGCGAGATCTGTGGTGGTCAATGAC 1039
Qy 1089 TCTGGCTTCGAGAGAGTACCAATGATTCAGAGCTGAGAGACACAGTCAATCAGTA 1148
Db 1040 TCTGGCTTCGAGAGAGTACCAATGATTCAGAGCTGAGAGACACAGTCAATCAGTA 1099
Qy 1149 TAAAAAGGAGAGAGCTTCGATATTAAGTTCGATCAGAGCTGAGAGTCAATGT 1208
Db 1100 TAAAAAGGAGAGAGCTTCGATATTAAGTTCGATCAGAGCTGAGAGTCAATGT 1159
Qy 1209 GCGATTTCTTTCTCTGCGCAAGTCTGGGCTGGGGTGCAGGCTGGGGCACTGCGCTCT 1268
Db 1160 GCGATTTCTTTCTCTGCGCAAGTCTGGGCTGGGGTGCAGGCTGGGGCACTGCGCTCT 1219
Qy 1269 GAGTCTGTCTGTGTTCTGGTTGCGCTGGCAATGTCTATCTGATGCTGGCTGTG 1328
Db 1220 GAGTCTGTCTGTGTTCTGGTTGCGCTGGCAATGTCTATCTGATGCTGGCTGTG 1279
Qy 1329 TCAATGCGCGGAAAGAACTACGCGGAGCTGGAGCATCTTTCAGCGCGGAAATCTACCA 1388
Db 1280 TCAATGCGCGGAAAGAACTACGCGGAGCTGGAGCATCTTTCAGCGCGGAAATCTACCA 1339
Qy 1389 TCCATATGAGAGTACCCACCTACCAACCCATGAGGCGCTATGTGCGCTACGAGTAC 1448
Db 1340 TCCATATGAGAGTACCCACCTACCAACCCATGAGGCGCTATGTGCGCTACGAGTAC 1399
Qy 1449 CGATGTCAGCGCTTATGAGAAAGTTTCTGCAAGTAAATGTGTGAGAGAGAGCTCTTAAC 1508
Db 1400 CGATGTCAGCGCTTATGAGAAAGTTTCTGCAAGTAAATGTGTGAGAGAGAGCTCTTAAC 1459
Qy 1509 AAACCCAGAGTGGCAGGCACTTCTGCGCACTTGTAGAGGAGCACTGCGCTCTGAGCTGA 1568
Db 1460 AAACCCAGAGTGGCAGGCACTTCTGCGCACTTGTAGAGGAGGAGCTGCGCTGAGCTGA 1519
Qy 1569 GTGG 1572
Db 1520 GTGG 1523
```

RESULT 8

```
A32135
LOCUS A32135 1452 bp mRNA linear PAT 15-JUL-1996
DEFINITION H.sapiens mammary carcinoma H23-ETA transmembrane antigen.
ACCESSION A32135
VERSION A32135.1 GI:1567303
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
location/Qualifiers
FEATURES
source
1..1452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="mammary carcinoma"
58..1425
/codon_start=1
/product="H23-ETA transmembrane antigen"
/protein_id="CAA02085.1"
/db_xref="GI:1567304"
/translation="MTPTQSPFFLLLLVLTIVTSGHASTPGEKETISATQRSS
VPSSTERNNAVSMTSVLSSHSPGSSSTTQGDVTLAPATEPASGAATWQDVTSVP
VTRPALGSTTPPAHDVTSAPDNKPAAGSTAPDXAHGVTSAPDXRPLAGSTAPVNVTS
ASGASGASASTLVHNGTSARATTPASKSTPPIPHSHSDTPTTASHSTYDASSTH
HSTVPLTSMNSTSPOLSTGVSPFPLSPHISLIPDNSSLEBDPDYIOELORDISEM
FLOIVKGGFLGISNKKRPGSVVOLLTAFRSGTINVDVETOFNOYKTEAARVYL
TISDVSYSVHPFRNSGAGYPMGIALILVLCVVALAIYVLIATLALVCCCRKRYNG
QLDIFPADTYPHMSRIFTHHGRIVPESSTDRSPLEKVSAGNGSSLSTNPVAA
TSANLV"
BASE COUNT 298 a 495 c 329 g 321 t 9 others
ORIGIN
Query Match 83.5%; Score 1313; DB 6; Length 1452;
Best Local Similarity 91.8%; Pred. No. 8,9e-230;
Matches 1443; Conservative 0; Mismatches 9; Indels 120; Gaps 1;
Qy 1 GAATTCCTGTGCTGTGAATCTGTTCGCCCCCTCCCACTTACCAACCACTG 60
Db 1 GAATTCCTGTGCTGTGAATCTGTTCGCCCCCTCCCACTTACCAACCACTG 60
Qy 61 ACACCGGAGCACCAAGTCCCTTCTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTT 120
Db 61 ACACCGGAGCACCAAGTCCCTTCTCTGCTGCTGCTCTCAAGTGTCTTACAGTTGTT 120
Qy 121 ACAGTTCTGTATGACAAAGCTTACCCAGGTGAGAAAGAGACTTGGCTTACCCAG 180
Db 121 ACAGTTCTGTATGACAAAGCTTACCCAGGTGAGAAAGAGACTTGGCTTACCCAG 180
Qy 181 AGAAGTTCAAGTCCAGCTCTACCTGAGAGAAATGCTGTGATGATGACACAGCTATTC 240
Db 181 AGAAGTTCAAGTCCAGCTCTACCTGAGAGAAATGCTGTGATGATGACACAGCTATTC 240
Qy 241 TCCAGGCAAGCGCCCGGTTCAAGGCTCTCCACCACTCAAGGAGACAGATGTCACTGAGCC 300
Db 241 TCCAGGCAAGCGCCCGGTTCAAGGCTCTCCACCACTCAAGGAGACAGATGTCACTGAGCC 300
Qy 301 CCGGCGAGGGAACCAAGCTTCAAGTTCAAGTGCACCTGGGAGACAGATGTCACTGAGTC 360
Db 301 CCGGCGAGGGAACCAAGCTTCAAGTTCAAGTGCACCTGGGAGACAGATGTCACTGAGTC 360
Qy 361 CCAATCACAGGCGACGCTGGGCTCCACCACTCCGCGACCAAGATGTCACTGAGTC 420
Db 361 CCAATCACAGGCGACGCTGGGCTCCACCACTCCGCGACCAAGATGTCACTGAGTC 420
Qy 421 CCGGACAAACAGACGCGCGGCTCCACCGCGCCCGCGGCGCAAGTGTCACTGAGTC 480
Db 421 CCGGACAAACAGACGCGCGGCTCCACCGCGCCCGCGGCGCAAGTGTCACTGAGTC 480
Qy 481 CCGGACAAACAGGCGCGCGGCTCCACCGCGCCCGCGGCGCAAGTGTCACTGAGTC 540
Db 481 CCGGACAAACAGGCGCGCGGCTCCACCGCGCCCGCGGCGCAAGTGTCACTGAGTC 540
Qy 541 CCGGACAAACAGGCGCGCGGCTCCACCGCGCCCGCGGCGCAAGTGTCACTGAGTC 600
```

| | | | | |
|---|---|------|--|------|
| D | b | 489 | ----- | 488 |
| Q | y | 601 | CGGAGCAACAGGCGGCGCTTGGGCTCCAGCGCCCTCCAGTCCAAATGTCAACTTCGGCC | 660 |
| D | b | 489 | -----NAGGCGANNNTTGGGCTCCACCGGCCCTCCAGTCCAAATGTCAACTTCGGCC | 540 |
| Q | y | 661 | TCAGGCTGTGCATCAGGCTCAGCTTCTACTCTGTGTCACACGGCACTCTGCGAGGCT | 720 |
| D | b | 541 | TCAGGCTGTGCATCAGGCTCAGCTTCTACTCTGTGTCACACGGCACTCTGCGAGGCT | 600 |
| Q | y | 721 | ACCAACAACCCGACGAGAAAGAGATCCACCCAGATTTCCAGCACAACCTGTACT | 780 |
| D | b | 601 | ACCAACAACCCGACGAGAAAGAGATCCACCCAGATTTCCAGCACAACCTGTACT | 660 |
| Q | y | 781 | CGTACCACTTGTCCAGCATAGACCAAGACTGATGCGAGTACACTCACATAGCACG | 840 |
| D | b | 661 | CGTACCACTTGTCCAGCATAGACCAAGACTGATGCGAGTACACTCACATAGCACG | 720 |
| Q | y | 841 | GTACCTCTCTCACTCCTCCCAATCAACAGACTTCCCGCAGTTGTCTACTGAGGCTCT | 900 |
| D | b | 721 | GTACCTCTCTCACTCCTCCCAATCAACAGACTTCCCGCAGTTGTCTACTGAGGCTCT | 780 |
| Q | y | 901 | TTCTTTTCTGTCTCTTTTCACATTTCAAACCTCCAGTTAATTCCTCTCGAAGATCC | 960 |
| D | b | 781 | TTCTTTTCTGTCTCTTTTCACATTTCAAACCTCCAGTTAATTCCTCTCGAAGATCC | 840 |
| Q | y | 961 | AGCAGCGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTAT | 1020 |
| D | b | 841 | AGCAGCGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTAT | 900 |
| Q | y | 1021 | AAACAAGGGGGGTTTTCTGGGCTCTCCCAATTTAAGTTCAAGCCAGATCTGTGGTGTA | 1080 |
| D | b | 901 | AAACAAGGGGGGTTTTCTGGGCTCTCCCAATTTAAGTTCAAGCCAGATCTGTGGTGTA | 960 |
| Q | y | 1081 | CAATTGACTCTTGACCTTTCGAGAAAGTACATCATATGTCCACAGCTGGAGACACATTC | 1140 |
| D | b | 961 | CAATTGACTCTTGACCTTTCGAGAAAGTACATCATATGTCCACAGCTGGAGACACATTC | 1020 |
| Q | y | 1141 | AATCACTATTAACCGAAGCAGCCTCTCATATAACCTGAAGATCTCAGACGTCAACGTG | 1200 |
| D | b | 1021 | AATCACTATTAACCGAAGCAGCCTCTCATATAACCTGAAGATCTCAGACGTCAACGTG | 1080 |
| Q | y | 1201 | AGTCATGTGCGCAATTTCTTTCTCTGCGCCAGCTCGAGGAGCGAGGCTGGGCGATC | 1260 |
| D | b | 1081 | AGTCATGTGCGCAATTTCTTTCTCTGCGCCAGCTCGAGGAGCGAGGCTGGGCGATC | 1140 |
| Q | y | 1261 | GCGCTGCTGGTGTGTGTGTCTGTGTTCTGGTTGCGCTGAGCCATTTGTATCTCATTTG | 1320 |
| D | b | 1141 | GCGCTGCTGGTGTGTGTGTCTGTGTTCTGGTTGCGCTGAGCCATTTGTATCTCATTTG | 1200 |
| Q | y | 1321 | GGTGTGTGTCACTGTGCGCGCGAAAGAACTACGAGGACAGTGGACATCTTTCAAGCCGAGAT | 1380 |
| D | b | 1201 | GGTGTGTGTCACTGTGCGCGCGAAAGAACTACGAGGACAGTGGACATCTTTCAAGCCGAGAT | 1260 |
| Q | y | 1381 | ACCTTACATCTCTATAGAGGAGTACCCCACTTACACACCCATGGGCGCTATGTGCGCCCT | 1440 |
| D | b | 1261 | ACCTTACATCTCTATAGAGGAGTACCCCACTTACACACCCATGGGCGCTATGTGCGCCCT | 1320 |
| Q | y | 1441 | AGCAGTACCGATCTGAGCCCTTATGAGAAAGTTTCTGAGGTAATGTGTGACAGACCTC | 1500 |
| D | b | 1321 | AGCAGTACCGATCTGAGCCCTTATGAGAAAGTTTCTGAGGTAATGTGTGACAGACCTC | 1380 |
| Q | y | 1501 | TCCTTACAAAACCCAGCAGTGGGACCACTTTCGCACTTTGTAGGGGACAGTGCGCCCT | 1560 |
| D | b | 1381 | TCCTTACAAAACCCAGCAGTGGGACCACTTTCGCACTTTGTAGGGGACAGTGCGCCCT | 1440 |
| Q | y | 1561 | TGAGCTGAGTGG 1572 | |
| D | b | 1441 | TGAGCTGAGTGG 1452 | |

| | | | |
|------------|---|---------|-----------------------------|
| AF423030 | | | |
| LOCUS | AF423030 | 1834 bp | mRNA linear SYN 11-MAR-2002 |
| DEFINITION | Synthetic construct Homo sapiens mucin variant MUC-1CT90 (MUC1) | | |
| ACCESSION | M8A, complete cds; alternatively spliced. | | |
| VERSION | AF423030 | | |
| KEYWORDS | AF423030.1 GI:19338619 | | |
| SOURCE | . | | |
| ORGANISM | synthetic construct. | | |
| TITLE | synthetic construct | | |
| REFERENCE | artificial sequences. | | |
| AUTHORS | 1 (bases 1 to 1834) Hinojosa-Kurtzberg,A.M., Johansson,M.J., Madsen,C.S., Hansson,G.C. and Gendler,S.J. | | |
| TITLE | Mucin MUC1 is part of the mucus accumulated in cystic fibrosis mice | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 1834) | | |
| AUTHORS | Hinojosa-Kurtzberg,A.M. and Gendler,S.J. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (21-SEP-2001) Biochemistry, Mayo Clinic, 13400 E. Shea Boulevard, Scottsdale, AZ 85259, USA | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..1834 | | |

| | |
|------|--------------|
| gene | 1. .1834 |
| | /gene="MUC1" |
| CDS | 1. .1452 |

/'note="alternatively spliced, contains exon 6a resulting in variant hydrophilic tail of carboxy-terminal domain; lacks sites for beta-catenin and Grb2 interactions; derived from Homo sapiens"

```

/locust_table=11
/trans_table=11
/product="mucin variant MUC-1CT80"
/protein_id="AAL86734.1"
/db_xref="GI:19338620"
/translaction="MPGCGSPFLILLITLVTVVSGSHASPNPGKEKTSATQRRSA
VPSSEKNAVSMSTSVLSHSPGSGSSTTGQDVTLATETPAGSAAITWQDVTSAV
VYRPLASTTTPPAHDVTSAPDNKPAPGSTAPPAHDSVADTPRAPSTAPPAHVSIT
ADPNKPSTAPPAVNTVTSASGASGATLNVHGTARATTPAPASKSTPTSIHSHH
STPTPLASTKSTKIDASSTHSTVPLTSSNSTPQSTSVFFFLSHLITLNPEN
SLEPSTVDQOELQDLSMFLOIYKGGEGFGLSNIKRPSVYVOLLTAAREGIN
HDTFGNRYKTEASRNMLTISVSVSDVPPFSAQSAVPEMGLITALLIVCTVAV
IAITVLAALVCCCRKNYQGLDLPARDITVHSEDTPTATHRVYVPSSTDRPYH
KRLPLGPGQCKYVWKGXDSGGTWTQRMKR"
```

ORIGIN

| Query Match | 80.7%; | Score 1269; | DB 12; | Length 1834; |
|-------------|--------|-------------|--------|--------------|
|-------------|--------|-------------|--------|--------------|

Matches 1345; Conservative 0; Mismatches 10; Indels 60; Gaps 1;

58 ATGACACCGGGCACCAGTCTCCTTCTTCTCCTGCTGCTCCTCCTCACAGTGTACAGTT 11

Db 1 ATGACACCGGACCCAGTCTCCTTCTTCCTGCTGCTCCTCACAAGTGTACAGTT 60

118 GTTACAGGTTCTGGTCATGCAAGCTTACCCGAGGTGGAGAAAGGAGACTTCGGCTACC 17

Db 61 GTTACAGGTTCTGGTCATGCAAGCTCTACCCAGGTGGAGAAAGGAGACTTCGGCTACC 12

QY 178 CAGAGAAGTTCAGTGCACCTCTACTGAGAGAATGCTGTGAGTAAGACCAGCAGCGTA 23

D6 121 CAGAGAAGTTCAGTGCCTCAGCTCTACCTGAGAGAAGATGCTGTGAGTATGACCAAGCAGCGTA 18

QY 238 CUCUCAGCCUACAGCCCTCCGTCUAGGTCUUCUCCACCACTCAGGACAGGATGTCACUCUG 29

DB 181 C T C T C C A G C C A C A G C C C C G T T C A G G C T C C T C C A C C A C T C A G G A C A G A G T C A C T C T G 24

298 GCCCUGGACUGAACUICIAGGICIAGCIGCCACCIUGGACAGGAGIACCCCG 33

RESULT 9

Db 241 GCCCGGCAACGAGCAAGCTTCAAGTTCACTGCGCACTGCGGAGAGATGCACTCG 300
 Qy 358 GTCCAGTCAACAGGCAAGCCCTGCGCTCAACACCCCGCAAGCCAGATGCACTCA 417
 Db 301 GTCCAGTCAACAGGCAAGCCCTGCGCTCAACACCCCGCAAGCCAGATGCACTCA 360
 Qy 418 GCCCGGCAACAGGCAAGCCCTGCGCTCAACACCCCGCAAGCCAGATGCACTCG 477
 Db 361 GCCCGGCAACAGGCAAGCCCGGCTCAACGCGCCCGCAAGCCAGATGCACTCG 402
 Qy 478 GCCCGGCAACAGGCGCGCGCGCGCTCAACGCGCCCGCAAGCCAGATGCACTCG 537
 Db 403 -----GCCACGATGCACTCG 420
 Qy 538 GCCCGGCAACAGGCGCGCGCGCTCAACGCGCCCGCAAGCCAGATGCACTCG 597
 Db 421 GCCCGGCAACAGGCGCGCGCGCTCAACGCGCCCGCAAGCCAGATGCACTCG 480
 Qy 598 GCCCGGCAACAGGCGCGCGCTCAACGCGCGCGCTCAACATGTCACCTCG 657
 Db 481 GCCCGGCAACAGGCGCGCGCTCAACGCGCGCGCTCAACATGTCACCTCG 540
 Qy 658 GCTCAGGCTCTGATCAGGCTCAGCTTCTACTGCTGTCACAAAGGCACTGCGCAG 717
 Db 541 GCTCAGGCTCTGATCAGGCTCAGCTTCTACTGCTGTCACAAAGGCACTGCGCAG 600
 Qy 718 GGTACCAACACCCAGCAAGGCAAGCACTCAACCCAGATTCGCGCAACCTCGAT 777
 Db 601 GGTACCAACACCCAGCAAGGCAAGCACTCAACCTCAATCCGCAACCTCGAT 660
 Qy 778 ACTCTACACCCCTGCGCAAGCAAGCACTGATGCGAGTACCTACCATAGC 837
 Db 661 ACTCTACACCCCTGCGCAAGCAAGCACTGATGCGAGTACCTACCATAGC 720
 Qy 838 AGGATACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGTGTCTAGGAGTC 897
 Db 721 AGGATACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGTGTCTAGGAGTC 780
 Qy 898 TCTTTCTTTTCTCTGCTTTTCACTTTCAAACTCGCAATTAATCTCTCTGGAAGAT 957
 Db 781 TCTTTCTTTTCTCTGCTTTTCACTTTCAAACTCGCAATTAATCTCTCTGGAAGAT 840
 Qy 958 CCCAGCAACGATCTACTACCAAGAGCTGCAAGAGACATTTCTGAATGTTTTGCAATT 1017
 Db 841 CCCAGCAACGATCTACTACCAAGAGCTGCAAGAGACATTTCTGAATGTTTTGCAATT 900
 Qy 1018 TATTAACAAAGGGGTTTTCTGGGCTCTCAATTAATTAAGTTGAGGAGATCTGAGTG 1077
 Db 901 TATTAACAAAGGGGTTTTCTGGGCTCTCAATTAATTAAGTTGAGGAGATCTGAGTG 960
 Qy 1078 GTACAAATTAAGTCTGAGCTTCCGAAAGAGTACATCAATGTCCACGACGTGAGACAG 1137
 Db 961 GTACAAATTAAGTCTGAGCTTCCGAAAGAGTACATCAATGTCCACGACGTGAGACAG 1020
 Qy 1138 TTCAATCGATTAATAAGAAAGAGAGCTCTGATATATACTGACAGCTCAGACCTGACG 1197
 Db 1021 TTCAATCGATTAATAAGAAAGAGAGAGCTCTGATATATACTGACAGCTCAGAC 1080
 Qy 1198 GTGAGTCAATGTCATTTCTTTCTCTGCGCAAGTCTGGGCTGCGAGGCTGGGGC 1257
 Db 1081 GTGAGTCAATGTCATTTCTTTCTCTGCGCAAGTCTGGGCTGCGAGGCTGGGGC 1140
 Qy 1258 ATCGGCTGCTGAGTCTGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
 Db 1141 ATCGGCTGCTGAGTCTGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 Qy 1318 TTGGCTGCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
 Db 1201 TTGGCTGCTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 Qy 1378 GATACCTCAACATCTATAGAGAGTACCCCACTTACCAACCCATGAGGCTATGTCGCC 1437
 Db 1261 GATACCTCAACATCTATAGAGAGTACCCCACTTACCAACCCATGAGGCTATGTCGCC 1320

Qy 1438 CTTAGCAGTACCGATGTAGCCCTATGAGAGGT 1472
 Db 1321 CTTAGCAGTACCGATGTAGCCCTATGAGAGGT 1355
 RESULT 10
 AF423031
 LOCUS
 DEFINITION
 mRNA, complete cds; alternatively spliced.
 ACCESSION
 AF423031
 VERSION
 AF423031
 KEYWORDS
 AF423031.1 GI:19338621
 SOURCE
 ORGANISM
 Hinojosa-Kurtzberg, A.M., Johanson, M.J., Madsen, C.S., Hansson, G.C. and Gendler, S.J.
 Mucin MUC1 is part of the mucus accumulated in cystic fibrosis mice
 REFERENCE
 1 (bases 1 to 1414)
 2 (bases 1 to 1414)
 Unpublished
 TITLE
 JOURNAL
 2 (bases 1 to 1414)
 Hinojosa-Kurtzberg, A.M. and Gendler, S.J.
 Direct Submission
 Submitted (21-SEP-2001) Biochemistry, Mayo Clinic, 13400 E. Shea
 Boulevard, Scottsdale, AZ 85259, USA
 FEATURES
 location/Qualifiers
 1..1414
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Homo sapiens gene in transgenic Mus musculus
 C57BL/6; isolated from intestinal mucosa"
 1..1414
 /gene="MUC1"
 1..1386
 /gene="MUC1"
 /note="alternatively spliced; contains exon 6b resulting
 in variant carboxy-terminal domain; lacks sites for
 beta-catenin and Gp2 interactions; derived from Homo
 sapiens"
 /codon_start=1
 /transl_table=1
 /product="mucin variant MUC1-CT58"
 /protein_id="AA186735.1"
 /db_xref="GI:19338621"
 /translation="MTPGQSPPELLLLITLVTVVSGHSSSTPGSEKTSATORS
 VSSSTERNAYKMSYSVLSRSGSSSTPGQDYLTAPEPESSAAATWQDVTSP
 VTRPALGSTTPPAHDVTSAPDNKRPAGSTAPPAAGVTSAPDTTPAPGSTAPPAHGTYS
 APDNRPALGSTAPPVHNVTSASGSASSTLVVNGTSAPATTPPAKSTPSPISRH
 SDPTPLASHSTKTDASHTSHSTVPPPLTSNHSSTPQSTGVSPFLSHSTISLDQFN
 SLEDSSTDYQDELORDISEMFLQIKQGGFLGLSNIKFRGSAVVGLTLAFEGEITNV
 HVEVTFQNYQKTEAASRYNLTIDVSGDVPFPSSAGSGAGVGMQIALIVAVCLVA
 LAIVLILAVCCQCRKNYGLDIFPARDYHPMSEVPTHTGRVVPSSSTDSRPE
 KEMRVDRERLA"
 BASE COUNT 299 a 481 c 335 g 299 t
 ORIGIN
 Query Match 80.7%; Score 1268; DB 12; Length 1414;
 Best Local Similarity 95.0%; Pred. No. 1,4e-221;
 Matches 1344; Conservative 0; Mismatches 10; Indels 60; Gaps 1;
 Qy 58 ATGACACCGGACCCAGCTCTCTTTCTTCTGCTGCTGCTCTCTCAAGTCTTACAGTT 117
 Db 1 ATGACACCGGACCCAGCTCTCTTTCTTCTGCTGCTGCTCTCTCAAGTCTTACAGTT 60
 Qy 118 GTTACAGGTTCTGATCAGAGCTTACCCAGGTGAGAAAGAGAGACTTCGGCTACC 177
 Db 61 GTTACAGGTTCTGATCAGAGCTTACCCAGGTGAGAAAGAGAGACTTCGGCTACC 120
 Qy 178 CAGAGAGTTCAAGTCCAGCTCTTCTGAGAAAGATGCTGTGATGATACAGCAGCTGA 237
 Db 121 CAGAGAGTTCAAGTCCAGCTCTTCTGAGAAAGATGCTGTGATGATACAGCAGCTGA 180

238 CTCTCAAGCCAGAGCCCGGGTTCAAGGCTCTCCACCACTCAAGGAGACAGATGTCACTCTG 297
Db 181 CTCTCAAGCCAGAGCCCGGGTTCAAGGCTCTCCACCACTCAAGGAGACAGATGTCACTCTG 240
Qy 298 GCCCGGCGCAAGGACCAAGCTTCAAGTTCAAGTTCACCTGGGAGACAGATGTCACTCTG 357
Db 241 GCCCGGCGCAAGGACCAAGCTTCAAGTTCAAGTTCACCTGGGAGACAGATGTCACTCTG 300
Qy 358 GTCCAGTCAACAGGCGAGCCCTGGGCTCCACACCCCGGCGAGCCCAAGATGTCACTCTCA 417
Db 301 GTCCAGTCAACAGGCGAGCCCTGGGCTCCACACCCCGGCGAGCCCAAGATGTCACTCTCA 360
Qy 418 GCCCGGCAACAGGCGAGCCCGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 477
Db 361 GCCCGGCAACAGGCGAGCCCGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 402
Qy 478 GCCCGGCAACAGGCGAGCCCGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 537
Db 403 -----GCCACGATGTCACTCTG 420
Qy 538 GCCCGGCAACAGGCGAGCCCGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 597
Db 421 GCCCGGCAACAGGCGAGCCCGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 480
Qy 598 GCCCGGCAACAGGCGAGCCCTGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 657
Db 481 GCCCGGCAACAGGCGAGCCCTGGGCTTCCACCCCGGCGAGCCCAAGATGTCACTCTG 540
Qy 658 GCGTCAAGGCTGTGATCAAGGCTCAAGTTCTACTGTGTGCAACAGGCACTCTGCGAGG 717
Db 541 GCGTCAAGGCTGTGATCAAGGCTCAAGTTCTACTGTGTGCAACAGGCACTCTGCGAGG 600
Qy 718 GGTACCAACACCCAGCGAGCAAGAGCACTCCACCAAGCAATTCGAGCCACCACTCTGAT 777
Db 601 GGTACCAACACCCAGCGAGCAAGAGCACTCCATTCATTCACAGCCACCACTCTGAT 660
Qy 778 ACTCTTACACCCCTTGGCGAGGCAAGACCAAGACTGATGCGAGTCACTCACTCACTAGC 837
Db 661 ACTCTTACACCCCTTGGCGAGGCAAGACCAAGACTGATGCGAGTCACTCACTCACTAGC 720
Qy 838 AGGTAACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGTTGTCTCACTGGGCTC 897
Db 721 AGGTAACCTCTCTCACTCTCTCAATCAAGCACTTCTCCAGTTGTCTCACTGGGCTC 780
Qy 898 TCTTCTTTTCTCTGTCTTCTTCAATTCACATTCACATTCAGATTTCTCTCTGAGAT 957
Db 781 TCTTCTTTTCTCTGTCTTCTTCAATTCACATTCAGATTTCTCTCTGAGAT 840
Qy 958 CCCAGCACCACTCACTCAAGAGCTGCAAGAGCAATTTCTGAATGTTTTTGCAGATT 1017
Db 841 CCCAGCACCACTCACTCAAGAGCTGCAAGAGCAATTTCTGAATGTTTTTGCAGATT 900
Qy 1018 TATTAACAAAGGGGGTTTTTCTGGGCTCTCCAAATTAAGTTCAAGGCGAGATGTGATG 1077
Db 901 TATTAACAAAGGGGGTTTTTCTGGGCTCTCCAAATTAAGTTCAAGGCGAGATGTGATG 960
Qy 1078 GTACAAATTAAGTCTGGGCTTCCAGAAAGTCAATGATGTCACAGCTGGAGACAAG 1137
Db 961 GTACAAATTAAGTCTGGGCTTCCAGAAAGTCAATGATGTCACAGCTGGAGACAAG 1020
Qy 1138 TTCAATCAAGTATTAACAAAGAGAGCTCTCCATTAACCTGAAGATCTCAAGCTCAAG 1197
Db 1021 TTCAATCAAGTATTAACAAAGAGAGCTCTCCATTAACCTGAAGATCTCAAGCTCAAG 1080
Qy 1198 GTGAGTCAATGATCTGCTTTCTCTTCTGAGCTGAGGCTGGGGTGCAGAGCTGGGGC 1257
Db 1081 GTGAGTCAATGATCTGCTTTCTCTTCTGAGCTGAGGCTGGGGTGCAGAGCTGGGGC 1140
Qy 1258 ATGAGGCTGAGTCAATGATCTGATGATCTGATGCTGAGCTGAGCTGATGCTCAATGCTC 1317
Db 1141 ATGAGGCTGAGTCAATGATCTGATGATCTGATGATGAGCTGAGCTGATGATGCTCAATGCTC 1200

Qy 1318 TTGAGCTGTCTGTCAATGTCGCCCGGAAAGAACTACCGGAGAGCTGAGACATCTTCCAGCCCG 1377
Db 1201 TTGAGCTGTCTGTCAATGTCGCCCGGAAAGAACTACCGGAGAGCTGAGACATCTTCCAGCCCG 1260
Qy 1378 GATACCTTACATCTTATGAGCGAGATACCCCACTTACCAACCCATGGGCGATATGTGCC 1437
Db 1261 GATACCTTACATCTTATGAGCGAGATACCCCACTTACCAACCCATGGGCGATATGTGCC 1320
Qy 1438 CCTAGCAATACCGATGTGAGCCCTTATGAGAGG 1471
Db 1321 CCTAGCAATACCGATGTGAGCCCTTATGAGAGG 1354

RESULT 11
AR030811
LOCUS AR030811 6192 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5861381.
ACCESSION AR030811
VERSION AR030811.1 GI:5944025
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 6192)
AUTHORS Chamhon,P., Kieny,M.-P., Iathe,R. and Hareuveni,M.
TITLES Pharmaceutical composition for the treatment of a malignant tumor
JOURNAL Patent: US 5861381-A 1 19-JAN-1999;
FEATURES
Location/Qualifiers
1..6192
source /organism="unknown"

BASE COUNT 695 a 2628 c 1513 g 636 t 720 others
ORIGIN

Query Match 76.1%; Score 1196.6; DB 6; Length 6192;
Best Local Similarity 92.2%; Pred. No. 1.5e-208;
Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 246 CCACAGCCCGGTTCAAGGCTCTCCACCACTCAAGGAGACAGATGTCACTCTGAGCCCGGC 305
Db 486 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 4925
Qy 306 CACGAAACCAAGTTCAAGTTCAAGTGCACCTGGGAGACAGATGTCACTCTGAGCCAGT 365
Db 4926 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 4985
Qy 366 CACAGGCCAGCCCTGGGCTTCCACCGCCCGCAGCCCAAGTGTCACTCTGAGCCCGGA 425
Db 4986 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 5045
Qy 426 CAACAGCCAGCCCGGCTTCCACCGCCCGCAGCCAGGTGTCACTCTGAGCCCGGA 485
Db 5046 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 5105
Qy 486 CACAGGCCGCCCCTGGGCTTCCACCGCCCGCAGCCAGGTGTCACTCTGAGCCCGGA 545
Db 5106 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 5165
Qy 546 CACAGGCCGCCCCTGGGCTTCCACCGCCCGCAGCCAGGTGTCACTCTGAGCCCGGA 605
Db 5166 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 5225
Qy 606 CAACAGGCCGCCCCTGGGCTTCCACCGCCCGCAGCCAGGTGTCACTCTGAGCCCGGA 665
Db 5226 CCNNAGGCCGNNCCGGGCTTCCACCGCCCGCANNNGCCACAGGTGTCACTCTGAGCCCGGA 5285
Qy 666 CTCTGATCAAGGCTCAAGCTTCTACTCTGTGTGCAACAGGCACTCTGAGCCAGGCTACAC 725
Db 5286 CTCTGATCAAGGCTCAAGCTTCTACTCTGTGTGCAACAGGCACTCTGAGCCAGGCTACAC 5345
Qy 726 AACCCAGCCAGAGAGCACTTCCACCAATTTCCAGCCAGCCAGCTGTGATCTCTTAC 785
Db 5346 AACCCAGCCAGAGAGCACTTCCACCAATTTCCAGCCAGCCAGCTGTGATCTCTTAC 5405

QY 786 CACCTTGGCAGCAGTACGACCAAGACTGATGCCAGTACGACTCACCATACAGGTTAC 845
 Db 5406 CACCTTGGCAGCAGTACGACCAAGACTGATGCCAGTACGACTCACCATACAGGTTAC 5465
 QY 846 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTCTTCTT 905
 Db 5466 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTCTTCTT 5525
 QY 906 TTTCTGTCCTTTTCAATTCAATTTCAATTCCTCTCTGAGAAATCCACAGAC 965
 Db 5526 TTTCTGTCCTTTTCAATTCAATTTCAATTCCTCTCTGAGAAATCCACAGAC 5585
 QY 966 CGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTATAACA 1025
 Db 5586 CGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTATAACA 5645
 QY 1026 AGGGGGTTTTCTGGGCTCTCCATATTAAATTGAGGCGAGATCTGGTGGTACATT 1085
 Db 5646 AGGGGGTTTTCTGGGCTCTCCATATTAAATTGAGGCGAGATCTGGTGGTACATT 5705
 QY 1086 GACTCTGGCTCTCCAGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 1145
 Db 5706 GACTCTGGCTCTCCAGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 5765
 QY 1146 GTATATAACGGAAGCAGCTCTCATATTAAGTACAGATCTCAGAGTTCAGAGTCA 1205
 Db 5766 GTATATAACGGAAGCAGCTCTCATATTAAGTACAGATCTCAGAGTTCAGAGTCA 5825
 QY 1206 TTGTCATTTCTCTTCTCTGCGCAGTCTGGGGCTGGGGTCCAGGCTGGGGATTCGGCT 1265
 Db 5826 TTGTCATTTCTCTTCTCTGCGCAGTCTGGGGCTGGGGTCCAGGCTGGGGATTCGGCT 5885
 QY 1266 GCTGTGTGCTGTGTCTGTGTCTGTGTGCGCTGGCATTGTCTATCTCATTTGGCTGT 1325
 Db 5886 GCTGTGTGCTGTGTCTGTGTCTGTGTGCGCTGGCATTGTCTATCTCATTTGGCTGT 5945
 QY 1326 CTGTGAGTGGCGCGAAAGACATAGAGGAGCTGGAATTTCCAGCCCGGAAATCACTA 1385
 Db 5946 CTGTGAGTGGCGCGAAAGACATAGAGGAGCTGGAATTTCCAGCCCGGAAATCACTA 6005
 QY 1386 CCATCTTAGAGAGTACCCCACTACCAACCCTATGGGCGCTATGTGCCCTTAGAC 1445
 Db 6006 CCATCTTAGAGAGTACCCCACTACCAACCCTATGTGGCGCTCTTAGAG 6065
 QY 1446 TACCGATGTAGCCCTATAGAGAGTTTCTGACAGTAATGTGTGAGAGGCTCTCTTA 1505
 Db 6066 TACCGATGTAGCCCTATAGAGAGTTTCTGACAGTAATGTGTGAGAGGCTCTCTTA 6125
 QY 1506 CACAAACCCAGAGTGGCAGCCACTTCTGCCAATTGTAGGGGCACTGGCCTTGAAC 1565
 Db 6126 CACAAACCCAGAGTGGCAGCCACTTCTGCCAATTGTAGGGGCACTGGCCTTGAAC 6185
 QY 1566 TGAGTGG 1572
 Db 6186 TGAGTGG 6192

RESULT 12
 ARI42537
 LOCUS ARI42537 6192 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 1 from patent US 6203795.
 ACCESSION ARI42537
 VERSION ARI42537.1 GI:15103823
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 6192)
 AUTHORS Chambon, F.; Kieny, M. P.; Iathe, R. and Hareuveni, M.
 TITLE Pharmaceutical Composition for the treatment or prevention of a
 malignant tumor
 JOURNAL Patent: US 6203795-A 1 20-MAR-2001;
 FEATURES Location/Qualifiers

source 1..6192
 BASE COUNT 695 a 2628 c 1513 g 636 t 720 others
 ORIGIN

Query Match 76.1%; Score 1196.6; DB 6; Length 6192;
 Best Local Similarity 92.2%; Pred. No. 1.5e-208;
 Matches 1223; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 246 CCACAGCCCGGTACAGCTCTCCACACTCAGAGCAGAGATGTCATCTGCGCCCGG 305
 Db 4866 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTCGGCGCGA 4925
 QY 306 CACGAAACAGCTTCAGTTCAAGTTCACCTGACCTGGAGACAGATGTCACTCGTCCAGT 365
 Db 4926 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTCGGCGCGA 4985
 QY 366 CACAGGCGAGCCCTGGGCTCCACACCGCCCGCAGGTCAGATGTCACTTCAAGCCCGA 425
 Db 4986 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTCGGCGCGA 5045
 QY 426 CAACAGCAGCTCCGGGCTCCACCGCCCGCAGGTCACCTTGGGCTCCGGA 485
 Db 5046 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTTGGGCTCCGGA 5105
 QY 486 CACAGGCGAGCCCGGCTCCACCGCCCGCAGGTCACCTTGGGCTCCGGA 545
 Db 5106 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTTGGGCTCCGGA 5165
 QY 546 CACAGGCGAGCCCGGCTCCACCGCCCGCAGGTCACCTTGGGCTCCGGA 605
 Db 5166 CANNAGGCGANNCCGGGCTCCACCGCCCGCANNNGCCACGGGTGACCTTGGGCTCCGGA 5225
 QY 606 CAACAGGCGAGCCCTGGGCTCCACCGCCCGCAGTCCAAATGTCACTTGGGCTCCAGG 665
 Db 5226 CANNAGGCGANNNTGGGCTCCACCGCCCGCAGTCCAAATGTCACTTGGGCTCCAGG 5285
 QY 666 CTCTGATCAGAGCTCAGCTTCTACTCTGTGTGCAAAAGGACCTGTGCAAGGCTACAC 725
 Db 5286 CTCTGATCAGAGCTCAGCTTCTACTCTGTGTGCAAAAGGACCTGTGCAAGGCTACAC 5345
 QY 726 AACCCAGCAGCAGAGACCTCCACCGAATTTCCAGCCACCACTGTATCTCTTAC 785
 Db 5346 AACCCAGCAGCAGAGACCTCCACCGAATTTCCAGCCACCACTGTATCTCTTAC 5405
 QY 786 CACCTTGGCAGCAGTACGACCAAGACTGATGCCAGTACACTCACTACAGGTAAC 845
 Db 5406 CACCTTGGCAGCAGTACGACCAAGACTGATGCCAGTACACTCACTACAGGTAAC 5465
 QY 846 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTTCTTCTT 905
 Db 5466 TCCTCTGACCTCTCCATACAGCACTTCTCCAGTTGTCTACTGGGGTCTTCTTCTT 5525
 QY 906 TTTCTGTCCTTTTCAATTCAATTTCAATTCCTCTCTGAGAAATCCACAGAC 965
 Db 5526 TTTCTGTCCTTTTCAATTCAATTTCAATTCCTCTCTGAGAAATCCACAGAC 5585
 QY 966 CGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTATAACA 1025
 Db 5586 CGACTACTACCAAGAGCTGACAGAGACATTTCTGAATGTTTTTGCAGATTTATAACA 5645
 QY 1026 AGGGGGTTTTCTGGGCTCTCCATATTAAATTGAGGCGAGATCTGGTGGTACATT 1085
 Db 5646 AGGGGGTTTTCTGGGCTCTCCATATTAAATTGAGGCGAGATCTGGTGGTACATT 5705
 QY 1086 GACTCTGGCTCTCCAGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 1145
 Db 5706 GACTCTGGCTCTCCAGAGAGTACATCATATGTCACAGCTGAGACACAGTTCAATCA 5765
 QY 1146 GTATATAACGGAAGCAGCTCTCATATTAAGTACAGATCTCAGAGTTCAGAGTCA 1205
 Db 5766 GTATATAACGGAAGCAGCTCTCATATTAAGTACAGATCTCAGAGTTCAGAGTCA 5825

[illegible]

| | | | |
|----|------|--|------|
| QY | 489 | CAGGCGCCCGGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 548 |
| Db | 2785 | CAGGCGGAGCCCGGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 2844 |
| QY | 549 | CAGGCGGAGCCCGGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 608 |
| Db | 2845 | CAGGCGGAGCCCGGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 2904 |
| QY | 609 | CAGGCGGAGCTTGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 668 |
| Db | 2905 | CAGGCGGAGCTTGGAGCTCCACCGCCCCCGCGCCACGGGTGTCACCTGCGGCCGGAGAC | 2964 |
| QY | 669 | TGATCAGGCTCAGCTTCTACTGTGTGTCACACGGGACTGTGCCAGGGCTTACCAAC | 728 |
| Db | 2965 | TGATCAGGCTCAGCTTCTACTGTGTGTCACACGGGACTGTGCCAGGGCTTACCAAC | 3024 |
| QY | 729 | CCGAGCGAGAAAGAGACTCACCCAGCAATGCCAGGACCACTGTGATCTCTACAC | 788 |
| Db | 3025 | CCGAGCGAGAAAGAGACTCACCTCAATTCCTCCAGGACCACTGTGATCTCTACAC | 3084 |
| QY | 789 | CTTTGCGAGCCATAGCACCAAGCTGATGCCAGTACCTCACCATAGCAGGTACTCC | 848 |
| Db | 3085 | CTTTGCGAGCCATAGCACCAAGCTGATGCCAGTACCTCACCATAGTGTGTGACTCC | 3144 |
| QY | 849 | TCTCAGCTCTCAGATACAGAGACTTCTCCAGGTGTCTACTGGGGCTCTTTCTTTT | 908 |
| Db | 3145 | TCTCAGCTCTCTCAGATACAGAGACTTCTCCAGGTGTCTACTGGGGCTCTTTCTTTT | 3204 |
| QY | 909 | CTGTCTTTTCATTTCAAACTCCAGTTTAATCTCTCTGGAAATCCGAGACCGA | 968 |
| Db | 3205 | CTGTCTTTTCATTTCAAACTCCAGTTTAATCTCTCTGGAAATCCGAGACCGA | 3264 |
| QY | 969 | CTACTACCAAGAGCTGAGAGAGACTTCTGAATGTTTTTGAGATTTATTAACAG | 1028 |
| Db | 3265 | CTACTACCAAGAGCTGAGAGAGACTTCTGAATGTTTTTGAGATTTATTAACAGG | 3324 |
| QY | 1029 | GGGTTTTCTGGGCTCTTCAATTAATTAAGTTCAGGCGAGAGTCTGTGTGTACATTAAC | 1088 |
| Db | 3325 | GGGTTTTCTGGGCTCTTCAATTAATTAAGTTCAGGCGAGAGTCTGTGTGTACAAATGAC | 3384 |
| QY | 1089 | TCTGAGCTTCCGAGAGAGTACATCATGTGCCAGCAGCTGGAGACACAGTTCAATCAGTA | 1148 |
| Db | 3385 | TCTGAGCTTCCGAGAGAGTACATCAATGTGCCAGCAGCTGGAGACACAGTTCAATCAGTA | 3444 |
| QY | 1149 | TAAAGCGAAGAGCCTCTGCATTAACCTGACAGATCTCAAGGTGAGGTGAGTACATGT | 1208 |
| Db | 3445 | TAAAGCGAAGAGCCTCTGCATTAACCTGACAGATCTCAAGGTGAGGTGAGTACATGT | 3504 |
| QY | 1209 | GCAATTTCTTTCTCTGCGCCAGTCTGGGGCTGGGGTGCAGGCTGGGGACATGCGCTGCT | 1268 |
| Db | 3505 | GCAATTTCTTTCTCTGCGCCAGTCTGGGGCTGGGGTGCAGGCTGGGGACATGCGCTGCT | 3564 |
| QY | 1269 | GATGCTGT | 1328 |
| Db | 3565 | GATGCTGT | 3624 |
| QY | 1329 | TCAATGCGCGCGAAGAACTACGCGGAGCTGACATCTTTCAGCCCGGAGTACTTACCA | 1388 |
| Db | 3625 | TCAATGCGCGCGAAGAACTACGCGGAGCTGACATCTTTCAGCCCGGAGTACTTACCA | 3684 |
| QY | 1389 | TCTCATAGAGAGTACCCCACTTACCAACCACTGGGGGCTATGTGCCCCCTTACGAGTAC | 1448 |
| Db | 3685 | TCTCATAGAGAGTACCCCACTTACCAACCACTGGGGGCTATGTGCCCCCTTACGAGTAC | 3744 |
| QY | 1449 | CGATCTGAGCCCTTATAGAAAGTTTTCTGAGGTTAAATGTGTGACAGCTCTTTTACAC | 1508 |
| Db | 3745 | CGATCTGAGCCCTTATAGAAAGTTTTCTGAGGTTAAATGTGTGACAGCTCTTTTACAC | 3804 |
| QY | 1509 | AAACCCAGAGTGGAGCACTTCGCAACTGTGTAGG | 1547 |
| Db | 3805 | AAACCCAGAGTGGAGCACTTCCTTCGCAACTGTGTAGG | 3843 |

RESULT 14
 AX335372 4139 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 5881 from Patent WO0194629.
 DEFINITION AX335372
 ACCESSION AX335372
 VERSION AX335372.1 GI:18126091
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
 Horigan, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 5881 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 source Location/Qualifiers
 1. 4139
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 632 a 1910 c 1055 g 542 t
 ORIGIN
 Query Match 75.7%; Score 1190.2; DB 6; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-207;
 Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 249 CAGCCCGGTTCCAGGCTCTCCACCACTCAGAGACAGAGATGTCATCTGCGCCCGCCAC 308
 DB 2545 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGCTGTCACCTCGGCCCGGACAC 2604
 QY 309 GGAAACAGCTTCAGGTTTCAGCTGCGACCTGGGAGACAGATGTCATCTCGGTCACAGTAC 368
 DB 2605 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGCTGTCACCTCGGCCCGGACAC 2664
 QY 369 CAGGCGGCGCTGGGCTCCACCGCCCGCCAGCCAGGATGTCATCTCGGCCCGGACAC 428
 DB 2665 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGATGTCATCTCGGCCCGGACAC 2724
 QY 429 CAAAGCAGCGCGCGGCTCCACCGCCCGCCAGGATGTCATCTCGGCCCGGACAC 488
 DB 2725 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGATGTCATCTCGGCCCGGACAC 2784
 QY 489 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGGATGTCATCTCGGCCCGGACAC 548
 DB 2785 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGATGTCATCTCGGCCCGGACAC 2844
 QY 549 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGGATGTCATCTCGGCCCGGACAC 608
 DB 2845 CAGGCGGCGCCCGGCGCTCCACCGCCCGCCAGCCAGGATGTCATCTCGGCCCGGACAC 2904
 QY 609 CAGGCGGCGCTGGGCTCCACCGCCCGCTCCAGTCCAGATGTCATCTCGGCCCGGACAC 668
 DB 2905 CAGGCGGCGCTGGGCTCCACCGCCCGCTCCAGTCCAGATGTCATCTCGGCCCGGACAC 2964
 QY 669 TCCATCAGGCTCAGCTTCTACTGTGTGCAACAAGCACTTTCGACGGGCTACCAAC 728
 DB 2965 TCCATCAGGCTCAGCTTCTACTGTGTGCAACAAGCACTTTCGACGGGCTACCAAC 3024
 QY 729 CCACAGCAGAAAGCACTCCACCCAGCATTCAGAGCACCACTTGATATCTTACAC 788
 DB 3025 CCACAGCAGAAAGCACTCCACCCAGCATTCAGAGCACCACTTGATATCTTACAC 3084
 QY 789 CCTTCAGGCGCATGACCAAGACTGATGCGAGTACCTCACAATAGACAGGATACCTCC 848
 DB 3085 CCTTCAGGCGCATGACCAAGACTGATGCGAGTACCTCACAATAGACAGGATACCTCC 3144
 QY 849 TCTCAGCTCTCCATCAGCAGCTTCTCCCAAGTGTCTACTGGGGCTCTTTCTTTT 908
 DB 3145 TCTCAGCTCTCCATCAGCAGCTTCTCCCAAGTGTCTACTGGGGCTCTTTCTTTT 3204

QY 909 CCTGTCTTTTACATTTCAAACTTCAGTTAAATTCCTCTCGAAAGATCCACACCGA 968
 DB 3205 CCTGTCTTTTACATTTCAAACTTCAGTTAAATTCCTCTCGAAAGATCCACACCGA 3264
 QY 969 CTATCAGCAGAGCTGCGAGAGACATTTTGAAATGTTTTTGAGATTATTAACAAG 1028
 DB 3265 CTATCAGCAGAGCTGCGAGAGACATTTTGAAATGTTTTTGAGATTATTAACAAG 3324
 QY 1029 GGGTTTTCTGGGCTCTCCAAATATTAGTTGAGCCAGATCTGTGTGTACATTGAC 1088
 DB 3325 GGGTTTTCTGGGCTCTCCAAATATTAGTTGAGCCAGATCTGTGTGTACATTGAC 3384
 QY 1089 TCTGGCTTCCAGAGAGTACCATCAATGTCCAGACGTGAGACACAGTTCAATCAGTA 1148
 DB 3385 TCTGGCTTCCAGAGAGTACCATCAATGTCCAGACGTGAGACACAGTTCAATCAGTA 3444
 QY 1149 TAAAGGAGAGAGCTCTCCGATAAAGTACGATCTGAGACGTGAGAGTACATGT 1208
 DB 3445 TAAAGGAGAGAGCTCTCCGATAAAGTACGATCTGAGACGTGAGAGTACATGT 3504
 QY 1209 GGCATTTCTTTTCTGCGCAAGTCTGGGCTGGGCTGCCAGCTGGGCAATCGGCTGCT 1268
 DB 3505 GGCATTTCTTTTCTGCGCAAGTCTGGGCTGGGCTGCCAGCTGGGCAATCGGCTGCT 3564
 QY 1269 GGTGCTGTCTGTTCTGTGGTGGGCGATGTCATCTCATTTGCTTGCTGTGCTG 1328
 DB 3565 GGTGCTGTCTGTTCTGTGGTGGGCGATGTCATCTCATTTGCTTGCTGTGCTG 3624
 QY 1329 TCAAGTCCCGGAAAGACTACGAGGAGCTGACATCTTTCAAGCCCGGATATCTGCA 1388
 DB 3625 TCAAGTCCCGGAAAGACTACGAGGAGCTGACATCTTTCAAGCCCGGATATCTGCA 3684
 QY 1389 TCTATGAGGAGTACCCCACTACACACCCATGAGGCGCTATGTCCTTACAGTAC 1448
 DB 3685 TCTATGAGGAGTACCCCACTACACACCCATGAGGCGCTATGTCCTTACAGTAC 3744
 QY 1449 CGATCGTAGCCCTATGAAAGTTTTCGAGGTAATGTGAGCAGACCTCTGTAC 1508
 DB 3745 CGATCGTAGCCCTATGAAAGTTTTCGAGGTAATGTGAGCAGACCTCTGTAC 3804
 QY 1509 AAACCCAGAGTGGAGCCTTTCGCAACTGTGAGG 1547
 DB 3805 AAACCCAGAGTGGAGCCTTTCGCAACTGTGAGG 3843
 RESULT 15
 AX336712 4139 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 7221 from Patent WO0194629.
 DEFINITION AX336712
 ACCESSION AX336712
 VERSION AX336712.1 GI:18127431
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
 Horigan, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 7221 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 source Location/Qualifiers
 1. 4139
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 632 a 1910 c 1055 g 542 t
 ORIGIN
 Query Match 75.7%; Score 1190.2; DB 6; Length 4139;
 Best Local Similarity 94.8%; Pred. No. 2.2e-207;

Matches 1231; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY 249 CAGCCCGGTTGAGGCTCTCTCCACCACTGAGGAGAGATGCTGGGCCCCGGGAC 308
Db 2545 CAGCCCGGCTCGGGCTTCCACCGCCCCCGAGCCAGGTTCTACTGGGCCCCGGGAC 2604
QY 309 GGAACGAGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 368
Db 2605 CAGGCGGGCCCCGGGCTTCCACCGCCCCCGAGCCAGGTTCTACTGGGCCCCGGGAC 2664
QY 369 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 428
Db 2665 CAGGCGGGCCCCGGGCTTCCACCGCCCCCGAGCCAGGTTCTACTGGGCCCCGGGAC 2724
QY 429 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 488
Db 2725 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 2784
QY 489 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 548
Db 2785 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 2844
QY 549 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 608
Db 2845 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 2904
QY 609 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 668
Db 2905 CAGGCGGCTTCAGTTTCAGTGTGCACTGGGGGAGAGATGTCTGCTGCTCCAGTCC 2964
QY 669 TGCATCAGGCTCAGCTTCTAATCTGTGTGCACAAGGACCTCTGCGAGGGCTTACCAAC 728
Db 2965 TGCATCAGGCTCAGCTTCTAATCTGTGTGCACAAGGACCTCTGCGAGGGCTTACCAAC 3024
QY 729 CCGACGAGCAAGAGCACTCCAGCATTTCCAGCCACCACTGTGATCTCTTACAC 788
Db 3025 CCGACGAGCAAGAGCACTCCAGCATTTCCAGCCACCACTGTGATCTCTTACAC 3084
QY 789 CCGTTCAGGCTTACGATTCGCAAGACTGATTCGCGATGACCTGACGATGACCGTACC 848
Db 3085 CCGTTCAGGCTTACGATTCGCAAGACTGATTCGCGATGACCTGACGATGACCGTACC 3144
QY 849 TCTGACCTCTCAATCAAGCACTTCTCCGAGTGTCTAATCTGGGGTCTCTTTT 908
Db 3145 TCTGACCTCTCAATCAAGCACTTCTCCGAGTGTCTAATCTGGGGTCTCTTTT 3204
QY 909 CCGTCTCTTTTCAATTTAAACCTCCAGTTTAACTCTCTGGAAGATCCGAGCACGA 968
Db 3205 CCGTCTCTTTTCAATTTAAACCTCCAGTTTAACTCTCTGGAAGATCCGAGCACGA 3264
QY 969 CTACTACCAAGAGCTGAGAGACATTTCTGAAATGTTTTGCAATTTATAACAAG 1028
Db 3265 CTACTACCAAGAGCTGAGAGACATTTCTGAAATGTTTTGCAATTTATAACAAG 3324
QY 1029 GGGTTTTCTGGGCTCTTCCAAATATTAAGTTGAGGCCAGGATCTGTGGTGGTACATTGAC 1088
Db 3325 GGGTTTTCTGGGCTCTTCCAAATATTAAGTTGAGGCCAGGATCTGTGGTGGTACATTGAC 3384
QY 1089 TCTGGGCTTCGAGAGAGTACATCAATGTCTCAAGAGCTGAGAGACAGTTCAATCAATA 1148
Db 3385 TCTGGGCTTCGAGAGAGTACATCAATGTCTCAAGAGCTGAGAGACAGTTCAATCAATA 3444
QY 1149 TAAAGAGAGAGAGCTCTGATTAATCAATGATCTGAGAGCTCAGGTCAGTCAATGAT 1208
Db 3445 TAAAGAGAGAGAGCTCTGATTAATCAATGATCTGAGAGCTCAGGTCAGTCAATGAT 3504
QY 1209 GGCATTTCTTTCTCTGCCCAGCTGGGGGCTGGGGGCTGCGGCTGCT 1268
Db 3505 GGCATTTCTTTCTCTGCCCAGCTGGGGGCTGGGGGCTGCGGCTGCT 3564
QY 1269 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1328
Db 3565 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3624

```

```

QY 1329 TCAGTGCCCGGAAAGAACTTACGGGCACTGGAGATTTTTCAGCCCGGAGTACCTTACCA 1388
Db 3625 TCAGTGCCTGGGAAAGAACTTACGGGCACTGGAGATTTTTCAGCCCGGAGTACCTTACCA 3684
QY 1389 TCCATGAGGAGATACCCCACTTACACACCCATGGGCGCTATGTGCCCCCTAGCAGTAC 1448
Db 3685 TCCATGAGGAGATACCCCACTTACACACCCATGGGCGCTATGTGCCCCCTAGCAGTAC 3744
QY 1449 CGATGTGACCCCTATGAGAGATTTTTCAGAGTAAATGTGAGGAGGAGGCTCTTACAC 1508
Db 3745 CGATGTGACCCCTATGAGAGATTTTTCAGAGTAAATGTGAGGAGGAGGCTCTTACAC 3804
QY 1509 AAACCCAGAGTGGAGCACTTCTGCGCACTTGTAGGG 1547
Db 3805 AAACCCAGAGTGGAGCACTTCTGCGCACTTGTAGGG 3843

```

Search completed: May 8, 2003, 16:17:28
 Job time : 4420 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2003, 22:59:10 ; Search time 1.16071 Seconds
(without alignments)
1033.205 Million cell updates/sec

Title: US-09-658-621B-26
Perfect score: 44
Sequence: 1 ISEMFLOIT 9

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 44 | 100.0 | 9 | 22 | AAU00563 |
| 2 | 44 | 100.0 | 10 | 22 | AAU00566 |
| 3 | 44 | 100.0 | 30 | 23 | AAU04995 |
| 4 | 44 | 100.0 | 216 | 21 | AAU71022 |
| 5 | 44 | 100.0 | 239 | 17 | AAU89423 |
| 6 | 44 | 100.0 | 240 | 17 | AAU89422 |
| 7 | 44 | 100.0 | 255 | 17 | AAU89420 |
| 8 | 44 | 100.0 | 264 | 17 | AAU89421 |
| 9 | 44 | 100.0 | 273 | 17 | AAU89418 |
| 10 | 44 | 100.0 | 282 | 17 | AAU89419 |

| | | | | | | |
|----|----|-------|------|----|----------|--------------------|
| 11 | 44 | 100.0 | 312 | 23 | AAU84810 | Human MUC1R consen |
| 12 | 44 | 100.0 | 321 | 21 | AAU84346 | Human cancer assoc |
| 13 | 44 | 100.0 | 327 | 16 | AAU86298 | Glycoprotein 39 C |
| 14 | 44 | 100.0 | 338 | 21 | AAU71028 | Ubiqutin-E. coli |
| 15 | 44 | 100.0 | 348 | 13 | AAU27662 | C-terminal region |
| 16 | 44 | 100.0 | 455 | 13 | AAU23973 | Transmembrane form |
| 17 | 44 | 100.0 | 455 | 21 | AAU71024 | Human Mucin 1 (MUC |
| 18 | 44 | 100.0 | 473 | 22 | AAU09508 | Human mucin-1 (MUC |
| 19 | 44 | 100.0 | 475 | 22 | AAU00573 | Human MUC1 polypep |
| 20 | 44 | 100.0 | 475 | 23 | AAU77476 | Human MUC1 seq ID |
| 21 | 44 | 100.0 | 495 | 22 | AAU00539 | Human MUC1 polypep |
| 22 | 44 | 100.0 | 577 | 21 | AAU71030 | Ubiqutin-E. coli |
| 23 | 44 | 100.0 | 5546 | 23 | AAU85008 | Human melanocyte d |
| 24 | 41 | 93.2 | 508 | 19 | AAU77233 | MiniMUC1 protein. |
| 25 | 36 | 81.8 | 1091 | 20 | AAU30337 | Protein encoded by |
| 26 | 33 | 75.0 | 57 | 21 | AAU60595 | Arabidopsis thalia |
| 27 | 33 | 75.0 | 61 | 21 | AAU60354 | Arabidopsis thalia |
| 28 | 33 | 75.0 | 71 | 21 | AAU47951 | Arabidopsis thalia |
| 29 | 33 | 75.0 | 1005 | 21 | AAU48223 | Arabidopsis thalia |
| 30 | 33 | 75.0 | 1057 | 21 | AAU48222 | Arabidopsis thalia |
| 31 | 33 | 75.0 | 1114 | 21 | AAU48221 | Arabidopsis thalia |
| 32 | 32 | 72.7 | 303 | 13 | AAU27654 | Human calcium chan |
| 33 | 32 | 72.7 | 1848 | 22 | AAU58215 | Drosophila melanog |
| 34 | 32 | 72.7 | 2262 | 19 | AAU56737 | Calcium ion channe |
| 35 | 32 | 72.7 | 2265 | 22 | AAU56702 | Rabbit P/O-type ca |
| 36 | 32 | 72.7 | 2266 | 16 | AAU71008 | Human neuronal cal |
| 37 | 32 | 72.7 | 2266 | 21 | AAU10580 | Human calcium chan |
| 38 | 32 | 72.7 | 2424 | 21 | AAU78901 | Calcium channel al |
| 39 | 32 | 72.7 | 2510 | 16 | AAU71007 | Human neuronal cal |
| 40 | 32 | 72.7 | 2510 | 21 | AAU10579 | Human calcium chan |
| 41 | 31 | 70.5 | 242 | 21 | AAU16708 | Bacteriophage Dp-1 |
| 42 | 31 | 70.5 | 522 | 22 | AAU46748 | R. marinus bacteri |
| 43 | 31 | 70.5 | 871 | 22 | AAU46727 | R. marinus bacteri |
| 44 | 31 | 70.5 | 1326 | 22 | AAU59665 | Drosophila melanog |
| 45 | 31 | 70.5 | 1639 | 19 | AAU54145 | P. falciparum synt |

ALIGNMENTS

| | | |
|----------|--|-----------------------------------|
| RESULT 1 | AAU00563 | AAU00563 standard; Peptide; 9 AA. |
| XX | AAU00563; | |
| XX | 12-SEP-2001 (first entry) | |
| XX | Human MUC1 polypeptide derivative #24. | |
| DE | Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I; | |
| KW | glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine; | |
| KW | cancer gene therapy; diagnosis; treatment; inflammatory disorder; | |
| KW | organ transplant rejection; graft versus host disease. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FN | WO200118035-A2. | |
| XX | | |
| PD | 15-MAR-2001. | |
| XX | | |
| PF | 07-SEP-2000; 2000WO-EP08761. | |
| XX | | |
| PR | 08-SEP-1999; 99GB-0021242. | |
| PR | 10-SEP-1999; 99EP-0402237. | |
| PR | 03-MAR-2000; 2000US-0187215. | |
| XX | | |
| PA | (TRGE) TRANSGENE SA. | |
| PA | (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD. | |
| XX | | |
| PI | Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CM; | |
| PI | Acres B, Thomas M; | |
| XX | | |

DR WPI: 2001-235187/24.
DR N-PSDB: AAS00609.
XX New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
XX
PS Claim 2, Page 74; 81pp; English.
XX The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumors.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
SQ Sequence 9 AA;
XX
XX Query Match 100.0%; Score 44; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ISEMPLOIY 9
|||
|||
|||
|||
|||
|||
|||
|||
|||
DB 1 ISEMPLOIY 9
|||
|||
|||
|||
|||
|||
|||
|||
|||
RESULT 2
AAU00566
XX AAU00566 standard; Peptide; 10 AA.
XX
AC AAU00566;
XX
XX 12-SEP-2001 (first entry)
XX
XX Human MUC1 polypeptide derivative #27.
XX
XX Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;
XX glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;
XX cancer gene therapy; diagnosis; treatment; inflammatory disorder;
XX organ transplant rejection; graft versus host disease.
XX
XX Homo sapiens.
XX
XX WO200118035-A2.
XX
XX 15-MAR-2001.
XX
XX 07-SEP-2000; 2000WO-EP08761.
XX
XX 08-SEP-1999; 99GB-0021242.
XX 10-SEP-1999; 99EP-0402237.
XX 03-MAR-2000; 2000US-0187215.
XX
XX (TRGE) TRANSGENE SA.
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX
XX Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CUM;
XX Acres B, Thomas M;
XX WPI: 2001-235187/24.
XX
XX

DR N-PSDB: AAS00712.
XX
XX New antigenic polypeptides of MUC-1 protein which activate cytotoxic T
PT lymphocyte proteins and their analogues, useful for identifying a major
PT histocompatibility complex class I restricted T cell response and for
PT diagnosing cancer -
XX
XX
PS Claim 2, Page 74; 81pp; English.
XX The sequence represents a human MUC1 polypeptide derivative. Derivative
CC antigenic peptides of MUC1 protein bind at least one major
CC histocompatibility complex class I (MHC-I) glycoprotein, which activates
CC cytotoxic T lymphocytes to induce a protective response against tumors.
CC Diagnosis of cancer involves determining the presence or absence in a
CC host cell of MHC class I restricted T cell response to a MUC1 derivative,
CC where the presence of the MHC class I restricted T cell response
CC indicates that the host has cancer. Measurement of the level of MHC class
CC I restricted T cell response is also useful to monitor the severity of
CC cancer, a larger response indicating a more severe cancer. MUC1
CC derivatives are useful in cancer therapy and to follow MUC1 specific
CC immune responses in patients during the course of disease and/or
CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and
CC diagnosis. Compositions of the sequences are used in vaccines and
CC treatments against cancer or diseases caused by an immune response, such
CC as an inflammatory disorder, organ transplant rejection or graft versus
CC host disease.
XX
SQ Sequence 10 AA;
XX
XX Query Match 100.0%; Score 44; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ISEMPLOIY 9
|||
|||
|||
|||
|||
|||
|||
|||
|||
DB 1 ISEMPLOIY 9
|||
|||
|||
|||
|||
|||
|||
|||
|||
RESULT 3
AAU84995
XX AAU84995 standard; Peptide; 30 AA.
XX
AC AAU84995;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human MUC1R segment 9.
XX
XX Svine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX viral infection; human immunodeficiency virus; melanoma;
XX bacterial infection; Salmonella; Legionella; parasitic infection;
XX Trypanosoma; Toxoplasma; Giardia.
XX
XX Homo sapiens.
XX
XX WO200190197-A1.
XX
XX 29-NOV-2001.
XX
XX 25-MAY-2001; 2001WO-AU00622.
XX
XX 26-MAY-2000; 2000AU-0007761.
XX
XX (AUSU) UNIV AUSTRALIAN NAT.
XX
XX Thomson SA, Ramshaw IA;
XX
XX WPI: 2002-147575/19.
XX N-PSDB: ABK36815.
XX
XX New synthetic polypeptides having several different segments of at
PT least one parent polypeptide linked together differently compared to
PT the linkage in the parent polypeptide, for inducing immune response
PT

PT against a pathogen or cancer -
XX
PS Example 3; Fig 27; 364pp; English.
XX
CC The invention relates to a new synthetic polypeptide (I) comprising
CC several different segments of at least one parent polypeptide linked
CC together in a different relationship relative to their linkage in the
CC parent polypeptide to impede, abrogate or otherwise alter at least one
CC function associated with the parent polypeptide and for inducing an
CC immune response against a pathogen or cancer. Also included are a
CC synthetic polynucleotide encoding and a computer system for
CC designing the synthetic polypeptides. The synthetic polypeptides and
CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
CC useful for modulating immune responses preferably directed against a
CC pathogen or a cancer (e.g., cancers of the lung, breast, ovary, cervix,
CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
CC Compositions comprising the polypeptide may be used in the treatment or
CC prophylaxis against viral (such as infections caused by HIV (human
CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
CC (e.g., infections caused by Neisseria, Meningococcus, Haemophilus,
CC Salmonella, Streptococcus, Legionella and Mycobacterium or parasitic
CC Trypanosoma, Toxoplasma and Giardia) infections. The present
CC sequence is a peptide derived from a parent protein used to
CC construct a vaccine of the invention.
XX
SQ Sequence 30 AA;
XX
Query Match 100.0%; Score 44; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ISEMFLOIY 9
12 ISEMFLOIY 20
DB
RESULT 4
AA71022
ID AA71022 standard; Protein; 216 AA.
XX
AC AA71022;
XX
DT 29-AUG-2000 (first entry)
XX
DE Human Mucin 1 (MUC-1) protein fragment #3.
XX
Human; Mucin 1; tumour; pMRS30 expression vector;
anti-tumour; therapy; immune response; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200025827-A2.
XX
PD 11-MAY-2000.
XX
PS 18-OCT-1999; 99WO-EP07874.
XX
PR 30-OCT-1998; 98IT-MT02330.
XX
PA (MENA) MENARINI RIGERCHE SPA.
PI Parente D, Di Massimo AM, De Santis R;
XX
XX WPI: 2000-365410/31.
DR N-PSDB; AAD00386.
XX
PT Composition containing one or more DNA molecules encoding fragments of
PT a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in
PT anti-tumor therapy -

PS Claim 16; Fig 3; 56pp; English.
XX
XX The present sequence is a fragment of human Mucin 1 (MUC-1), an
CC antigenic protein overexpressed in tumour cells. The sequence was
CC obtained from BT20 tumour cells. The corresponding DNA sequence
CC is cloned into a pMRS30 expression vector and used in pharmaceutical
CC composition e.g. vaccine for inducing an antigen-specific anti-tumour
CC immune response. Composition containing this DNA molecule is useful in
CC anti-tumour therapy of patients affected with tumours characterised
CC by high MUC-1 expression.
XX
SQ Sequence 216 AA;
XX
Query Match 100.0%; Score 44; DB 21; Length 216;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ISEMFLOIY 9
108 ISEMFLOIY 116
DB
RESULT 5
AAR89423
ID AAR89423 standard; Protein; 239 AA.
XX
AC AAR89423;
XX
DT 24-APR-1996 (first entry)
XX
DE Mucin-derived protein MUC1/V/alt.
XX
XX MUC1; MUC1/X/alt; MUC1/Y/alt; MUC1/V/alt; MUC1/V/alt;
XX MUC1/W/alt; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.
XX
OS Homo sapiens.
XX
PN WO9603502-A2.
XX
PD 08-FEB-1996.
XX
PF 21-JUL-1995; 95WO-IB00627.
XX
PR 26-JUL-1994; 94IL-0110464.
XX
PA (UPRA-) UNIV RAMOT APPL RES & IND DEV LTD.
XX
XX Wreschner DH;
XX
XX WPI: 1996-117047/12.
DR N-PSDB; AAT10682.
XX
PT Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
PT cancer
XX
PS Claim 6; Fig 6D; 79pp; English.
XX
XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
CC MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
CC differentially expressed in human breast cancer tissue and significantly
CC enhance the in vivo tumorigenic potential of mammary epithelial cells.
CC They serve as cell surface receptor molecules participating in signal
CC transduction. The proteins can be obtd. by expression of encoding cDNA
CC (see AAT10677-82) in recombinant host cells. They are used in the
CC treatment of human breast cancer and as diagnostic reagents. Receptor
CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
XX isolated.
XX
SQ Sequence 239 AA;
XX
Query Match 100.0%; Score 44; DB 17; Length 239;

Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 57 ISEMFLQIY 65

RESULT 6

AA89422 ID AAR89422 standard; Protein: 240 AA.

AC AAR89422;

DT 24-APR-1996 (first entry)

DE Mucin-derived protein MUC1/V.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
receptor; diagnosis; imaging; therapy.

XX Homo sapiens.

XX WO9603502-A2.

XX 08-FEB-1996.

XX 21-JUL-1995; 95WO-IB00627.

XX 26-JUL-1994; 94IL-0110464.

XX (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

XX Wreschner DH;

XX WPI; 1996-117047/12.

XX N-PSDB; AAT10681.

XX Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
PT cancer

XX Claim 6; Fig 6C; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
differentially expressed in human breast cancer tissue and significantly
enhance the in vivo tumorigenic potential of mammary epithelial cells.
XX They serve as cell surface receptor molecules participating in signal
transduction. The proteins can be obtd. by expression of encoding cDNA
CC (see AAT10677-82) in recombinant host cells. They are used in the
CC treatment of human breast cancer and as diagnostic reagents. Receptor
CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
CC isolated.

XX Sequence 240 AA;

Query Match 100.0%; Score 44; DB 17; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.34; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 58 ISEMFLQIY 66

RESULT 7

AA89420 ID AAR89420 standard; Protein: 255 AA.

XX AAR89420;

XX

DT 24-APR-1996 (first entry)

XX Mucin-derived protein MUC1/Y.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.

XX Homo sapiens.

XX WO9603502-A2.

XX 08-FEB-1996.

XX 21-JUL-1995; 95WO-IB00627.

XX 26-JUL-1994; 94IL-0110464.

XX (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD.

XX Wreschner DH;

XX WPI; 1996-117047/12.

XX N-PSDB; AAT10679.

XX Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt,
PT Z, Z/alt - for the diagnosis, imaging and therapy of human breast
PT cancer

XX Claim 5; Fig 6A; 79pp; English.

XX Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt,
MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and
differentially expressed in human breast cancer tissue and significantly
enhance the in vivo tumorigenic potential of mammary epithelial cells.
XX They serve as cell surface receptor molecules participating in signal
transduction. The proteins can be obtd. by expression of encoding cDNA
CC (see AAT10677-82) in recombinant host cells. They are used in the
CC treatment of human breast cancer and as diagnostic reagents. Receptor
CC ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been
CC isolated.

XX Sequence 255 AA;

Query Match 100.0%; Score 44; DB 17; Length 255;
Best Local Similarity 100.0%; Pred. No. 0.36; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISEMFLQIY 9
Db 73 ISEMFLQIY 81

RESULT 8

AA89421 ID AAR89421 standard; Protein: 264 AA.

XX AAR89421;

XX 24-APR-1996 (first entry)

XX Mucin-derived protein MUC1/Y/alt.

XX MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt;
XX MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer;
XX receptor; diagnosis; imaging; therapy.

XX Homo sapiens.

XX WO9603502-A2.

XX 08-FEB-1996.

XX

| | | |
|----------|--|-------------------------------------|
| PF | 21-JUL-1995; | 95WO-IB00627. |
| XX | | |
| PR | 26-JUL-1994; | 94IL-0110464. |
| XX | | |
| PA | (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD. | |
| XX | | |
| PI | Wreschner DH; | |
| XX | | |
| DR | UPI; 1996-117047/12. | |
| DR | N-PsDB; AAT10680. | |
| XX | | |
| PT | Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt, | |
| PT | Z, Z/alt - for the diagnosis, imaging and therapy of human breast | |
| PT | cancer | |
| PS | Claim 5, Fig 6B; 79pp; English. | |
| XX | | |
| CC | Novel, phosphorylated mucin-derived proteins MUC1/X, MUC1/X/alt, | |
| CC | MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt (AAR89418-23) are highly and | |
| CC | differentially expressed in human breast cancer tissue and significantly | |
| CC | enhance the in vivo tumorigenic potential of mammary epithelial cells. | |
| CC | They serve as cell surface receptor molecules participating in signal | |
| CC | transduction. The proteins can be obtd. by expression of encoding cDNA | |
| CC | (see AAT10677-82) in recombinant host cells. They are used in the | |
| CC | treatment of human breast cancer and as diagnostic reagents. Receptor | |
| CC | ligands MUC1/W, W/alt, Z and Z/alt (AAR89424-27) have also been | |
| CC | isolated. | |
| XX | | |
| SO | Sequence 264 AA; | |
| | | |
| | Query Match 100.0%; Score 44; DB 17; Length 264; | |
| | Best Local Similarity 100.0%; Pred. No. 0.37; | |
| | Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| OY | 1 ISEMFLOIY 9 | |
| | | |
| Db | 82 ISEMFLOIY 90 | |
| | | |
| RESULT 9 | | |
| AAR89418 | ID | AAR89418 standard; Protein; 273 AA. |
| XX | | |
| AC | AAR89418; | |
| XX | | |
| DT | 24-APR-1996 (first entry) | |
| XX | | |
| DE | Mucin-derived protein MUC1/X. | |
| XX | | |
| KM | MUC1; MUC1/X; MUC1/X/alt; MUC1/Y; MUC1/Y/alt; MUC1/V; MUC1/V/alt; | |
| KM | MUC1/W; MUC1/W/alt; MUC1/Z; MUC1/Z/alt; mucin; breast cancer; | |
| KM | receptor; diagnosis; imaging; therapy. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | WO9603502-A2. | |
| XX | | |
| PD | 08-FEB-1996. | |
| XX | | |
| PF | 21-JUL-1995; 95WO-IB00627. | |
| XX | | |
| PR | 26-JUL-1994; 94IL-0110464. | |
| XX | | |
| PA | (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD. | |
| XX | | |
| PI | Wreschner DH; | |
| XX | | |
| DR | UPI; 1996-117047/12. | |
| DR | N-PsDB; AAT10677. | |
| XX | | |
| PT | Mucin derived proteins MUC1/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt, | |
| PT | Z, Z/alt - for the diagnosis, imaging and therapy of human breast | |
| PT | cancer | |

| | | |
|-----------|--|--|
| XX | PS | Claim 4; Fig 5A; 79pp; English. |
| XX | CC | Novel, phosphorylated mucin-derived proteins MUCl/X, MUCl/X/alt, |
| XX | CC | MUCl/Y, MUCl/Y/alt, MUCl/V and MUCl/V/alt (AA89418-23) are highly and |
| CC | CC | differentially expressed in human breast cancer tissue and significantly |
| CC | CC | enhance the in vivo tumorigenic potential of mammary epithelial cells. |
| CC | CC | They serve as cell surface receptor molecules participating in signal |
| CC | CC | transduction. The proteins can be obt'd. by expression of encoding cDNA |
| CC | CC | (see ARI0677-82) in recombinant host cells. They are used in the |
| CC | CC | treatment of human breast cancer and as diagnostic reagents. Receptor |
| CC | CC | ligands MUCl/W, W/alt, Z and Z/alt (AA89424-27) have also been |
| CC | CC | isolated. |
| XX | SQ | Sequence 273 AA; |
| XX | Query Match | 100.0%; Score 44; DB 17; Length 273; |
| XX | Best Local Similarity | 100.0%; Pred. No. 0.39; |
| XX | Matches | 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| OY | 1 ISMFLOIY 9 | ' |
| | | |
| Dd | 91 ISMFLOIY 99 | |
| RESULT 10 | | |
| AA89419 | ID | AA89419 standard; Protein; 282 AA. |
| XX | AA89419; | |
| AC | 24-APR-1996 | (first entry) |
| XX | DE | Mucin-derived protein MUCl/X/alt. |
| XX | KX | MUC1: MUCl/X, MUCl/X/alt; MUCl/Y, MUCl/Y/alt; MUCl/V, MUCl/V/alt; |
| XX | KW | MUCl/W, MUCl/W/alt; MUCl/Z, MUCl/Z/alt; mucin; breast cancer; |
| KM | receptor; diagnosis; imaging; therapy. | |
| XX | OS | Homo sapiens. |
| XX | PN | MO603502-A2. |
| XX | PD | 08-FEB-1996. |
| XX | PF | 21-JUL-1995; 95WO-IB00627. |
| XX | PR | 26-JUL-1994; 94IL-0110464. |
| XX | PA | (UYRA-) UNIV RAMOT APPL RES & IND DEV LTD. |
| XX | PI | Wreschner DH; |
| XX | DR | WFI; 1996-117047/12. |
| XX | N-PSDB; | AAT10678. |
| XX | PT | Mucin derived proteins MUCl/X, X/alt, Y, Y/alt, V, V/alt, W, W/alt, |
| XX | PT | Z, Z/alt - for the diagnosis, imaging and therapy of human breast |
| XX | PT | cancer |
| XX | PS | Claim 4; Fig 5B; 79pp; English. |
| XX | CC | Novel, phosphorylated mucin-derived proteins MUCl/X, MUCl/X/alt, |
| XX | CC | MUCl/Y, MUCl/Y/alt, MUCl/V and MUCl/V/alt (AA89418-23) are highly and |
| CC | CC | differentially expressed in human breast cancer tissue and significantly |
| CC | CC | enhance the in vivo tumorigenic potential of mammary epithelial cells. |
| CC | CC | They serve as cell surface receptor molecules participating in signal |
| CC | CC | transduction. The proteins can be obt'd. by expression of encoding cDNA |
| CC | CC | (see ARI0677-82) in recombinant host cells. They are used in the |
| CC | CC | treatment of human breast cancer and as diagnostic reagents. Receptor |
| CC | CC | ligands MUCl/W, W/alt, Z and Z/alt (AA89424-27) have also been |
| CC | CC | isolated. |

SQ Sequence 282 AA;
 Query Match 100.0%; Score 44; DB 17; Length 282;
 Best Local Similarity 100.0%; Pred. No. 0.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMFLOIY 9
 |||||
 Db 100 ISEMFLOIY 108

RESULT 11
 AAB4810
 ID AAB4810 standard; Protein; 312 AA.
 XX
 AC AAB4810;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human WC1R consensus sequence.
 XX
 KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficiency virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.
 XX
 OS Homo sapiens.
 XX
 PN WO200190197-A1.
 XX
 PD 29-NOV-2001.
 XX
 PF 25-MAY-2001; 2001WO-AU00622.
 XX
 PR 26-MAY-2000; 2000AU-000761.
 XX
 PA (AUSU) UNIV AUSTRALIAN NAT.
 XX
 PI Thomson SA, Ramshaw IA;
 XX
 WP1; 2002-147575/19.
 XX
 PT New synthetic polypeptides having several different segments of at
 PT least one parent polypeptide linked together differently compared to
 PT the linkage in the parent polypeptide, for inducing immune response
 PT against a pathogen or cancer
 XX
 PS Example 3; Fig 27; 364pp; English.
 XX
 CC The invention relates to a new synthetic polypeptide (I) comprising
 CC several different segments of at least one parent polypeptide linked
 CC together in a different relationship relative to their linkage in the
 CC parent polypeptide to impede, abrogate or otherwise alter at least one
 CC function associated with the parent polypeptide and for inducing an
 CC immune response against a pathogen or cancer. Also included are a
 CC synthetic polynucleotide encoding and a computer system for
 CC designing the synthetic polypeptides. The synthetic polypeptides and
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is
 CC useful for modulating immune responses preferably directed against a
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present
 CC sequence is a consensus sequence for a parent protein used to design a
 CC Savine of the invention.
 XX

SQ Sequence 312 AA;
 Query Match 100.0%; Score 44; DB 23; Length 312;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ISEMFLOIY 9
 |||||
 Db 130 ISEMFLOIY 138

RESULT 12
 AAB43416
 ID AAB43416 standard; Protein; 321 AA.
 XX
 AC AAB43416;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated protein sequence SEQ ID NO:861.
 XX
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vunerary; immunomodulator;
 KW antidiabetic; antiaesthetic; antineumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nocrotic;
 KW vasotropic; antiprotatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 XX
 OS Homo sapiens.
 XX
 PN WO200055550-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCT INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WP1; 2000-58753/55.
 XX
 DR N-PSDB; AAC77625.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX
 PS Claim 11; Page 1414-1416; 2352pp; English.
 XX
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vunerary; immunomodulator;
 CC antidiabetic; antiaesthetic; antineumatic; antiarthritic;
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC nocrotic; vasotropic; antiprotatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to

CC AACT8457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 321 AA;

Query Match 100.0%; Score 44; DB 21; Length 321;
 Best Local Similarity 100.0%; Pred. No. 0.46;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 139 ISEMFLQIY 147

RESULT 13
 AAR6298 *
 ID AAR6298 standard; Peptide; 327 AA.
 XX
 AC AAR6298;

DT 26-JUL-1996 (first entry)

DE Glycoprotein 39 C terminal fragment.

XX Glycoprotein 39; gp39; lambda gII1 cDNA library; gastric cancer;
 KW cell line KATO-III; tumour; immune abnormality; marker;
 KW inflammatory disease.

OS Homo sapiens.

PN JP07051065-A.

PD 28-FEB-1995.

PF 21-FEB-1992; 92JP-0035085.

PR 21-FEB-1992; 92JP-0035085.

PA (NIKO-) NIPPON KOTAI KENKYUSHO KK.
 XX (UYKA-) UNIV KAGOSHIMA.

DR WPI; 1995-127356/17.

DR N-PSDB; AAT29190.

PT Glyco:protein 39 gene - used in the mass production of glyco:protein
 PT 39, for use as tumour and immune abnormality marker

PS Claim 2; Page 9-10; 10pp; Japanese.

CC The sequences given in AAR6297-98 represent portions of glycoprotein
 CC 39. This sequence represents the C terminal portion of the glycoprotein
 CC 39 gene. The cDNA encoding this sequence was isolated from lambda gII1
 CC cDNA library derived from the gastric cancer cell line KATO-III.
 CC Glycoprotein 39 is expected to be used as a tumour marker, an immune
 CC abnormality marker or a marker for various inflammatory diseases.

SQ Sequence 327 AA;

Query Match 100.0%; Score 44; DB 16; Length 327;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 145 ISEMFLQIY 153

RESULT 14
 AAY71028
 ID AAY71028 standard; Protein; 338 AA.
 XX
 AC AAY71028;

DT 29-AUG-2000 (first entry)

XX Ubiquitin-E. coli LacI-human Mucin 1 fusion protein #3.

XX Ubiquitin; LacI; beta-galactosidase; fusion protein;

XX human; Mucin 1; MUC-1; tumour; pMR30 expression vector;

XX anti-tumour; therapy; immune response; cytostatic; vaccine.

OS Chimeric - Homo sapiens.

XX Chimeric - Escherichia coli.

FT Key Location/Qualifiers

FT Region 1..123 /label= UBILacI_protein

FT /note= "contains ubiquitin sequence fused to a

FT portion of E. coli LacI"

FT Region 124..338 /label= Human_MUC-1_fragment

XX WO200025827-A2.

XX 11-MAY-2000.

XX 18-OCT-1999; 99WO-EP07874.

XX 30-OCT-1998; 98IT-M102330.

XX (MENA) MENARINI RICERCHE SPA.

XX Parente D, Di Massimo AM, De Santis R;

XX WPI; 2000-365410/31.

XX N-PSDB; AAD00392.

XX Composition containing one or more DNA molecules encoding fragments of

XX a Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in

XX anti-tumour therapy.

XX Claim 18; Fig 9; 56pp; English.

XX The present sequence is a fusion protein consisting

XX of human Mucin 1 (MUC-1) fragment fused to UBILacI sequence at the

XX N-terminus. The UBILacI sequence consists of ubiquitin from MCF7 cell

XX line and a portion of E. coli beta-galactosidase (LacI). MUC-1 is an

XX antigenic protein overexpressed in tumour cells. The corresponding

XX cDNA sequence is cloned into a pMR30 expression vector and used in

XX pharmaceutical composition e.g. vaccine for inducing an antigen-specific

XX anti-tumour immune response. Composition containing this DNA molecule

XX characterised by high MUC-1 expression.

SQ Sequence 338 AA;

Query Match 100.0%; Score 44; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 0.48;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISEMFLQIY 9
 |||||
 DB 230 ISEMFLQIY 238

RESULT 15
 AAR27662
 ID AAR27662 standard; Protein; 348 AA.
 XX
 AC AAR27662;

DT 06-NOV-1992 (first entry)

DE C-terminal region of H23-ETA-T antigen.

XX Transmembrane; human epithelial antigen; Monoclonal antibody H23;

